

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2002, 14:17:37 ; Search time 37.71 Seconds
(without alignments)
3550.731 Million cell updates/sec

Title: US-09-803-165-34
Perfect score: 4026
Sequence: 1 MFDYIRKDKPIRIRK.....KEDLKYSQKQGLDAMLKK 774

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_ARCHAEA:*
2: SP_BACTERIA:*
3: SP_FUNGI:*
4: SP_HUMAN:*
5: SP_INVERTEBRATE:*
6: SP_MAMMAL:*
7: SP_MHC:*
8: SP_ORGANELLE:*
9: SP_PHAGE:*
10: SP_PLANT:*
11: SP_PROTOZOA:*
12: SP_VIRUS:*
13: SP_UNCLASSIFIED:*
14: SP_VIRUS:*
15: SP_VIRUS:*
16: SP_BACTERIA:*
17: SP_ARCHAEA:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3212	79.8	775	1 09P9K4	09P9K4 pyrococcus
2	3211	79.8	775	1 09H06	09H06 pyrococcus
3	3081	76.5	759	1 09H98	09H98 pyrococcus
4	1204.5	29.9	781	1 09P9N1	09P9N1 sulfurispha
5	1204.5	29.9	781	1 09P9Y1	09P9Y1 sulfolobus
6	1190	29.6	803	1 059691	059691 pyrococcus
7	1125	27.9	785	1 09P9M2	09P9M2 pyrococcus
8	774	19.2	195	1 09H85	09H85 thermococcus
9	631.5	15.7	1443	5 09NAH1	09NAH1 caenorhabdi
10	626	15.5	901	17 09HRV9	09HRV9 halobacteri
11	623	15.5	796	17 09HJRO	09HJRO halobacteri
12	622.5	15.5	800	17 097AH3	097AH3 thermoplasm
13	616.5	15.3	876	17 0971C7	0971C7 sulfolobus
14	616.5	15.3	1105	11 091YV0	091YV0 mus musculu
15	616	15.3	1107	4 096H98	096H98 homo sapien
16	584.5	14.5	914	1 059690	059690 pyrococcus

17	567	14.1	1492	10 09FNA3	09FNA3 arabidopsis
18	559.5	13.9	223	17 026310	026310 methanother
19	555.5	13.8	1458	13 09DE46	09DE46 xenopus lae
20	527	13.1	1472	5 09VD90	09VD90 drosophila
21	516.5	12.8	844	1 031096	031096 cenarchaeum
22	509	12.6	845	1 074046	074046 cenarchaeum
23	507.5	12.6	1001	12 090626	090626 porcine lym
24	506.5	12.6	1004	12 090628	090628 porcine lym
25	506.5	12.6	1415	5 077034	077034 drosophila
26	504.5	12.5	2154	10 092VC8	092VC8 arabidopsis
27	491	12.2	913	12 09Y229	09Y229 chlorella v
28	488	12.2	2207	3 093845	093845 porcine cyt
29	487	12.1	1007	12 091F13	091F13 porcine cyt
30	486	12.1	1007	12 091F15	091F15 porcine cyt
31	486	12.1	1007	12 091F16	091F16 porcine cyt
32	484	12.0	1855	5 09TX75	09TX75 plasmodium
33	484	12.0	1855	5 09BHD0	09BHD0 plasmodium
34	483	12.0	2271	10 09SGD5	09SGD5 arabidopsis
35	481	11.9	1007	12 091F17	091F17 porcine cyt
36	481	11.9	1009	12 09YMQ4	09YMQ4 ateline her
37	481	11.9	1912	5 09U0H1	09U0H1 plasmodium
38	479	11.9	1016	12 0993K6	0993K6 callitrichi
39	478.5	11.9	1013	12 092827	092827 human herpe
40	477.5	11.9	774	2 09F175	09F175 pseudomonas
41	477	11.8	2243	4 09UNF3	09UNF3 homo sapien
42	477	11.8	2286	4 09Y554	09Y554 homo sapien
43	477	11.8	2297	4 09Y555	09Y555 homo sapien
44	471.5	11.7	1041	12 09DKT8	09DKT8 elephant he
45	468.5	11.6	1012	12 040910	040910 kaposi's sa

ALIGNMENTS

RESULT 1
ID 09P9K4 PRELIMINARY: PRT: 775 AA.
AC 09P9K4:
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE DNA POLYMERASE (EC 2.7.7.7).
GN DNA POL.
OS Pyrococcus glycoformans.
CC Archaea; Euryarchaeota; Thermococcales; Pyrococcus.
OX NCBI_TaxID=74610;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL585;
RA Querellon J., Cambon M.A., Lesongeur F., Forterre P., Barbier G.;
RT "DNA polymerase genes organisation of species belonging to
Thermococcales and phylogenetic implications".
RT Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N
PYROPHOSPHATE + DNA(N).
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC EMBL: AJ004834; CAB81809.1; -
DR HSSP; P56689; JNCO.
DR InterPro; IPR002064; DNA_POL_B.
DR Pfam; PF00136; DNA_POL_B; 1.
DR Pfam; PF03104; DNA_POL_B_exo; 1.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLB; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; UNKNOWN.1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase;
Nucleic acid transferase; Transferase.
SQ SEQUENCE 775 AA; 90492 MW; 677264920D770C0F CRC64;

Query Match 79.8%; Score 3212; DB 1; Length 775;
Best Local Similarity 77.1%; Pred. No. 1.1e-180;
Matches 595; Conservative 86; Mismatches 89; Indels 2; Gaps 1;

```

QY 1 MFDOTYITKDGKPIIRIFKKEGEFFKIELDPHPQPIYALALDKDSALDEIKAKIGERRHG 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MILDADYITEDGKPIIRIFKKEGEFFKIEVDNRFRPIYALALDKDSQIDVKKITAEHRHG 60
QY 61 KIYRVVDAAVKKKFLGRDVEYKKLFEHPQDYPALRGKIREHPAYIDIEYDIPFAKRY 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KIYRIYDVEYKKKFLGRPLEYKKLFEHPQDYPALRGKIREHPAYIDIEYDIPFAKRY 120
QY 121 LIDKGLIPMEGDEELKLMADIEFFYHEGDEFKGEIIMISYADEEARIVITKKNIDLPY 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIDKGLIPMEGDEELKLMADIEFFYHEGDEFKAGPTIMSYADEEGAKIYTKKKYDLPY 180
QY 181 VDVAVSNEREMIKRFPQIVREKDPDLITVNGDNFDPYLKRAEKLGVTLLGRDKEHPE 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VEYVSSEREMIKRFPQIVREKDPDLITVNGDSFDPYLKRAEKLGKIGKPLGRDGS--E 238
QY 241 PKIHRMGDSFAVEIKGRIHDLFPVVRRTINLPTYLEAVEYEAIVLCKTSKLAEEIAAI 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 PKMORLGDMTAVEIKGRIHDLVHVRRTINLPTYLEAVEYEAIVFGPKKRYAHEIAEA 298
QY 301 WETEESMKLAQYSMEDARATYELGKEFFPMEAEALKIGOSVADYSRSSTGNLVEYLL 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 WETGKGLERAKYSMDAKTYELGREFPMEANQLSLVGQPLMDYSRSSTGNLVEYLL 358
QY 361 RAVARNELAPNKPDEBEYRRRLRTTYLGIVKPEPERGMENTIYDFRCLYSIIVTHN 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 RKAVERNELAPNKPDEBEYRRRLRESYAGGYVKEPEKGLVSDFRSLYPSIIITHN 418
QY 421 VSPDTLEREGCKNYDVAPIYGVKFCDFPFIISIGELITMROEIKKKKATIDPFEK 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 VSPDTLNRREGCKMEYDVAPEYKHKFCDFPFIISLKLRLDROEIKRKKAKSDPLEK 478
QY 481 MLDYRORAVKLANSYGYGVKPKARWYSKCAESYATAMGRHYIEMTIKEIEKFGFKVL 540
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 MLDYRORAIKILANSYGYGVKAKARWYKCAESYATAMGREYIEFVRKLEKFGFKVL 538
QY 541 YADTDFYATIPGKPEPTIKKAKKEFLKYINSKLPGLLEVEYGFYLRGFFVAKKRYAVI 600
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 YIDTDLGATIPGAKPEEIKRKALEVEYINAKLPGLLEVEYGFYVRGFFVAKKRYALI 598
QY 601 DEEGRIITTRGLEVVRDWSIAKETQAKVLEALIKEDSVKAVEIYKDVVEELAKYQVPL 660
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 DEEGKIITTRGLEVVRDWSIAKETQAKVLEALIKGNVEAVKIYKEVTEKISKEIIP 658
QY 661 EKLVIHEQITKDSSEYKALGPVAAIKRLAKIKVPRGTIISYIVLRGSGKISDRYILL 720
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 EKLVIHEQITRPLHEYKALGPVAAIKRLAAGVYKVRPGVAVIGYIVLRGSGPISKRAIIA 718
QY 721 SEYDPRKKHKYDPDYIENQVLPVAVLILAEFGYRKEDLKYQSSQVGLDAML 772
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 719 EEPDPRKKHKYDAEYIENQVLPVAVLILAEFGYRKEDLKMOKTKQVGLTAML 770

```

```

RESULT 2
09HH06 PRELIMINARY: PRT: 775 AA.
AC 09HH06:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DNA POLYMERASE.
GN POL.
OS Pyrococcus glycovorans.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=74610;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AL646;
RA Querrelou J.J.B., Cambon M.A., Lesongeur F., Barbier G.;
RT "Thermococcales taxonomy and phylogeny based on the comparative use of
  16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
  genes".
  Submitted (OCR-1999) to the EMBL/GenBank/DBJ databases.

```

```

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N
CC PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: AJ250335; CAC12849.1; -.
DR HSSP: P56689; 1TGO.
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam: PF00136; DNA_pol_B; 1.
DR PRINTS: PR00106; DNAPOB.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN_1.
KW DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ SEQUENCE 775 AA; 90417 MW; 633F8392EBEBC73 CRC64;

```

```

Query Match 79.8%; Score 3211; DB 1; Length 775;
Best local similarity 77.1%; Pred. No. 12e-180;
Matches 595; Conservative 86; Mismatches 89; Indels 2; Gaps 1;

```

```

QY 1 MFDOTYITKDGKPIIRIFKKEGEFFKIELDPHPQPIYALALDKDSALDEIKAKIGERRHG 60
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MILDADYITEDGKPIIRIFKKEGEFFKIEVDNRFRPIYALALDKDSQIDVKKITAEHRHG 60
QY 61 KIYRVVDAAVKKKFLGRDVEYKKLFEHPQDYPALRGKIREHPAYIDIEYDIPFAKRY 120
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KIYRIYDVEYKKKFLGRPLEYKKLFEHPQDYPALRGKIREHPAYIDIEYDIPFAKRY 120
QY 121 LIDKGLIPMEGDEELKLMADIEFFYHEGDEFKGEIIMISYADEEARIVITKKNIDLPY 180
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIDKGLIPMEGDEELKLMADIEFFYHEGDEFKAGPTIMSYADEEGAKIYTKKKYDLPY 180
QY 181 VDVAVSNEREMIKRFPQIVREKDPDLITVNGDNFDPYLKRAEKLGVTLLGRDKEHPE 240
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VEYVSSEREMIKRFPQIVREKDPDLITVNGDSFDPYLKRAEKLGKIGKPLGRDGS--E 238
QY 241 PKIHRMGDSFAVEIKGRIHDLFPVVRRTINLPTYLEAVEYEAIVLCKTSKLAEEIAAI 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 PKMORLGDMTAVEIKGRIHDLVHVRRTINLPTYLEAVEYEAIVFGPKKRYAHEIAEA 298
QY 301 WETEESMKLAQYSMEDARATYELGKEFFPMEAEALKIGOSVADYSRSSTGNLVEYLL 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 WETGKGLERAKYSMDAKTYELGREFPMEANQLSLVGQPLMDYSRSSTGNLVEYLL 358
QY 361 RAVARNELAPNKPDEBEYRRRLRTTYLGIVKPEPERGMENTIYDFRCLYSIIVTHN 420
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 359 RKAVERNELAPNKPDEBEYRRRLRESYAGGYVKEPEKGLVSDFRSLYPSIIITHN 418
QY 421 VSPDTLEREGCKNYDVAPIYGVKFCDFPFIISIGELITMROEIKKKKATIDPFEK 480
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 419 VSPDTLNRREGCKMEYDVAPEYKHKFCDFPFIISLKLRLDROEIKRKKAKSDPLEK 478
QY 481 MLDYRORAVKLANSYGYGVKPKARWYSKCAESYATAMGRHYIEMTIKEIEKFGFKVL 540
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 MLDYRORAIKILANSYGYGVKAKARWYKCAESYATAMGREYIEFVRKLEKFGFKVL 538
QY 541 YADTDFYATIPGKPEPTIKKAKKEFLKYINSKLPGLLEVEYGFYLRGFFVAKKRYAVI 600
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 YIDTDLGATIPGAKPEEIKRKALEVEYINAKLPGLLEVEYGFYVRGFFVAKKRYALI 598
QY 601 DEEGRIITTRGLEVVRDWSIAKETQAKVLEALIKEDSVKAVEIYKDVVEELAKYQVPL 660
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 599 DEEGKIITTRGLEVVRDWSIAKETQAKVLEALIKGNVEAVKIYKEVTEKISKEIIP 658
QY 661 EKLVIHEQITKDSSEYKALGPVAAIKRLAKIKVPRGTIISYIVLRGSGKISDRYILL 720
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 EKLVIHEQITRPLHEYKALGPVAAIKRLAAGVYKVRPGVAVIGYIVLRGSGPISKRAIIA 718
QY 721 SEYDPRKKHKYDPDYIENQVLPVAVLILAEFGYRKEDLKYQSSQVGLDAML 772
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 719 EEPDPRKKHKYDAEYIENQVLPVAVLILAEFGYRKEDLKMOKTKQVGLTAML 770

```

RESULT 3

PS Disclosure: Fig 9: 40pp: English.

CC The present invention relates to thermostable mutant B-type DNA
CC polymerases, which have a Y-GG/A amino acid motif between the N-terminal
CC 3'-5' exonuclease domain and the C-terminal polymerase domain, where the
CC lysine of this motif is mutated. The mutant B-type DNA polymerase is
CC useful for synthesizing nucleic acids and for PCR. The present sequence
CC is the protein sequence for a recombinant *Thermococcus* aggreghans (Tag)
CC DNA polymerase, which was used to illustrate the invention.

XX Sequence 774 AA:

Query Match 100.0%; Score 4026; DB 22; Length 774;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 774; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFDNDYTRKDGKPIIRIRKKEGFEKILDPHOPYITALLKDSADDEIKAKGERHG 60
1 mfdndytrkdgkpiirirrkkegfekildphopiyitalldksaddeikakgerhg 60
DB 61 KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
DB 121 LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
QY 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
DB 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
QY 241 PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
DB 241 PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
DB 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
QY 361 RAYERNELAPNKPDEEYRRRLRTYLGYYKEPERGMENTYIDFCLYPSIIVTN 420
RAYERNELAPNKPDEEYRRRLRTYLGYYKEPERGMENTYIDFCLYPSIIVTN 420
DB 361 RAYERNELAPNKPDEEYRRRLRTYLGYYKEPERGMENTYIDFCLYPSIIVTN 420
RAYERNELAPNKPDEEYRRRLRTYLGYYKEPERGMENTYIDFCLYPSIIVTN 420
QY 421 VSPDTEREGCKNYDAPIVGYKCFKDPGFIPIISIGELITMBOELKTKMKATIDPIEK 480
VSPDTEREGCKNYDAPIVGYKCFKDPGFIPIISIGELITMBOELKTKMKATIDPIEK 480
DB 421 VSPDTEREGCKNYDAPIVGYKCFKDPGFIPIISIGELITMBOELKTKMKATIDPIEK 480
VSPDTEREGCKNYDAPIVGYKCFKDPGFIPIISIGELITMBOELKTKMKATIDPIEK 480
QY 481 MLDYRGAVKLHANSYYGYWGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKEGFKVL 540
MLDYRGAVKLHANSYYGYWGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKEGFKVL 540
DB 481 MLDYRGAVKLHANSYYGYWGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKEGFKVL 540
MLDYRGAVKLHANSYYGYWGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKEGFKVL 540
QY 541 YATDGFYATIPGEKPEETIKKAKKEFLKYTNKPLLELYEGFYLGRGFVAKRYAVI 600
YATDGFYATIPGEKPEETIKKAKKEFLKYTNKPLLELYEGFYLGRGFVAKRYAVI 600
DB 541 YATDGFYATIPGEKPEETIKKAKKEFLKYTNKPLLELYEGFYLGRGFVAKRYAVI 600
YATDGFYATIPGEKPEETIKKAKKEFLKYTNKPLLELYEGFYLGRGFVAKRYAVI 600
QY 601 DEGRITTTGLEVVRDMSIEAKETOAKVLEALIKEDSVKAEVIEIKVDVEIETAKQVPL 660
DEGRITTTGLEVVRDMSIEAKETOAKVLEALIKEDSVKAEVIEIKVDVEIETAKQVPL 660
DB 601 DEGRITTTGLEVVRDMSIEAKETOAKVLEALIKEDSVKAEVIEIKVDVEIETAKQVPL 660
DEGRITTTGLEVVRDMSIEAKETOAKVLEALIKEDSVKAEVIEIKVDVEIETAKQVPL 660
QY 661 EKLIVHEQITKDLSEKKAIGPHVAIAKRLAANGIKVRPGTIIISYIVLRSGKISDRVILL 720
EKLIVHEQITKDLSEKKAIGPHVAIAKRLAANGIKVRPGTIIISYIVLRSGKISDRVILL 720
DB 661 EKLIVHEQITKDLSEKKAIGPHVAIAKRLAANGIKVRPGTIIISYIVLRSGKISDRVILL 720
EKLIVHEQITKDLSEKKAIGPHVAIAKRLAANGIKVRPGTIIISYIVLRSGKISDRVILL 720
QY 721 SEVDPRKKHYDPRDYTENOVLPVLRILLEFGYRKEDLVQSSKOGLDAMLK 774
SEVDPRKKHYDPRDYTENOVLPVLRILLEFGYRKEDLVQSSKOGLDAMLK 774
DB 721 SEVDPRKKHYDPRDYTENOVLPVLRILLEFGYRKEDLVQSSKOGLDAMLK 774
SEVDPRKKHYDPRDYTENOVLPVLRILLEFGYRKEDLVQSSKOGLDAMLK 774

RESULT 2
AAW29323
XD AAW29323 standard: Protein: 774 AA.

XX AAW29323;
AC 20-APR-1998 (first entry)
XX
DT DNA polymerase with 3'-5' exonuclease activity.
XX
DE
XX
XX
KW TY Exon; DSM 10597; thermostable; DNA polymerase;
XX 3'-5' exonuclease; amplification.
OS *Thermococcus* sp.
XX
XX DE19611759-A1.
XX
PD 02-OCT-1997.
XX
XX
PF 25-MAR-1996; 96DE-1011759.
XX
XX 25-MAR-1996; 96DE-1011759.
XX
XX (BOEF) BOEHRINGER MANNHEIM GMBH.
XX
XX Antrenkian G, Frey B, Niehaus F;
XX WPI: 1997-48194/45.
XX N-PSDB: AAT86434.
XX
PT Thermostable DNA polymerase from *Thermococcus* sp. TY - useful for
XX nucleic acid amplification
PS Claim 1: Pages 9-10; 32pp; German.
XX
XX The present sequence (TY Exon) is a *Thermococcus* sp. TY
XX (DSM 10597) thermostable DNA polymerase with 3'-5' exonuclease
XX activity.
XX The enzyme can specifically amplify nucleic acid fragments of up to
XX 5 kb in high yields, has an activity half-life of 20 minutes at 90
XX degrees C, has an optimum temperature of 70-80 degrees C, has an
XX optimum pH of 7.5, exhibits optimum activity at a KCl concentration
XX of 80-100 mM, its magnesium ion-dependent and is inhibited by
XX manganese ions.
XX
XX
SQ Sequence 774 AA:

Query Match 99.1%; Score 3990; DB 18; Length 774;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 769; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MFDNDYTRKDGKPIIRIRKKEGFEKILDPHOPYITALLKDSADDEIKAKGERHG 60
1 mfdndytrkdgkpiirirrkkegfekildphopiyitalldksaddeikakgerhg 60
DB 1 mfdndytrkdgkpiirirrkkegfekildphopiyitalldksaddeikakgerhg 60
mfdndytrkdgkpiirirrkkegfekildphopiyitalldksaddeikakgerhg 60
QY 61 KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
DB 61 KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
KIVRVDAVKVKKKFLGRVDEWKLFEEHPODPAALRGKIREHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
DB 121 LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
LIDKGLIPMEGDEELKLMFADIEFTYHEGDEFGKEIIMISYADEBARVITWKNIDLPY 180
QY 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
DB 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKIGVLLGRDKEHP 240
QY 241 PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
DB 241 PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
PKIHRGDSFAVEIKRIFHDLFPVVRPTINLPTYLEAVYEAVLSEKTSKLGAEETIAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
DB 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360
WETESMKKLAQYSMEDARATYELGKEFFPMEAEKLIGQSVWDVSRSTGNLVEMYL 360

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2002, 14:17:02 ; Search time 34.46 Seconds

(without alignments)
2494.809 Million cell updates/sec

Title: US-09-803-165-34

Percent score: 4026
Sequence: 1 MFRTDITRKDKPKPIIRIFK.....KEDLYQSSKOYGLDAMLKK 774

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Hit number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /Geneseq_032802.*
2: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
3: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
4: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
5: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
6: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
7: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
8: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
9: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
10: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
11: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
12: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
13: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
14: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
15: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
16: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
17: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
18: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
19: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
20: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
21: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
22: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
23: /SIDSL/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4026	100.0	774	22	Recombinant Tag DN
2	3990	99.1	774	18	DNA polymerase wit
3	3446.5	85.6	1829	18	DNA polymerase wit
4	3356	83.4	1702	14	T. litoralis DNA p
5	3232.5	80.3	773	19	Thermococcus gorgo
6	3231.5	80.3	773	19	Thermococcus gorgo
7	3226.5	80.1	774	18	Thermococcus pepto
8	3225.5	80.1	774	19	Heat-resistant DNA
9	3222.5	80.0	774	17	Mature DNA polymer
10	3222.5	80.0	774	18	KOD1 DNA polymer
11	3221.5	80.0	774	19	Mutant KOD DNA pol

12	3220.5	80.0	774	19	AAW48452	Mutant KOD DNA pol
13	3220.5	80.0	774	19	AAW48458	Mutant KOD DNA pol
14	3219.5	80.0	774	19	AAW48450	Mutant KOD DNA pol
15	3218.5	79.9	774	19	AAW48454	Mutant KOD DNA pol
16	3218.5	79.9	774	19	AAW48455	Mutant KOD DNA pol
17	3218.5	79.9	774	19	AAW48457	Mutant KOD DNA pol
18	3218.5	79.9	774	19	AAW48459	Mutant KOD DNA pol
19	3218.5	79.9	774	19	AAW48460	Mutant KOD DNA pol
20	3218.5	79.9	774	19	AAW48461	Mutant KOD DNA pol
21	3217.5	79.9	774	19	AAW48462	Mutant KOD DNA pol
22	3215.5	79.9	774	18	AAW33112	Thermococcus pepto
23	3213.5	79.8	774	19	AAW44731	T. profundus therm
24	3213.5	79.8	774	19	AAW44732	Mutant KOD DNA pol
25	3211.5	79.8	774	19	AAW48451	Thermococcus DNA p
26	3203.5	79.6	776	22	AAU03085	Thermococcus DNA p
27	3202.5	79.5	776	22	AAU03075	Thermococcus DNA p
28	3202.5	79.5	776	22	AAU03084	Thermococcus DNA p
29	3199.5	79.5	776	22	AAU03086	Thermococcus DNA p
30	3199.5	79.5	776	22	AAU03088	Thermococcus DNA p
31	3199.5	79.5	776	22	AAU03090	Thermococcus DNA p
32	3198.5	79.4	776	22	AAU03081	Thermococcus DNA p
33	3198.5	79.4	776	22	AAU03083	Thermococcus DNA p
34	3198.5	79.4	776	22	AAU03087	Thermococcus DNA p
35	3197.5	79.4	776	22	AAU03076	Thermococcus DNA p
36	3196.5	79.4	776	22	AAU03079	Thermococcus DNA p
37	3196.5	79.4	776	22	AAU03089	Thermococcus DNA p
38	3195.5	79.4	776	22	AAU03078	Thermococcus DNA p
39	3195.5	79.4	776	22	AAU03080	Thermococcus DNA p
40	3195.5	79.4	776	22	AAU03091	Thermococcus DNA p
41	3195.5	79.4	776	22	AAU03092	Thermococcus DNA p
42	3193.5	79.3	776	20	AAW87548	Thermococcus DNA p
43	3192.5	79.3	776	22	AAU03077	Thermococcus DNA p
44	3164.5	78.6	771	22	AAW91133	Thermococcus abyssi
45	3164.5	78.6	774	22	AAW96414	putative P. abyssi

ALIGNMENTS

RESULT 1
ID AAG78941 standard; Protein: 774 AA.
AC AAG78941;
XX 13-FEB-2002 (first entry)
DT Recombinant Tag DNA polymerase.
XX Tag polymerase; B-type DNA polymerase.
KW Tag polymerase; B-type DNA polymerase.
XX Thermococcus aggregans.
OS Thermococcus aggregans.
PN EPI132474-A1.
XX 13-SEP-2001.
PD 13-SEP-2001.
XX 06-MAR-2001; 2001EP-0104583.
PF 11-MAR-2000; 2000EP-0105155.
PR (HOFF) ROCHE DIAGNOSTICS GMBH.
PA Sobek H, Frey B, Antranikian G, Boehlke K, Pilsani FM, Rossi M;
PI Sobek H, Frey B, Antranikian G, Boehlke K, Pilsani FM, Rossi M;
XX WPI: 2001-618367/72.
DR N-PSDB: AAT70053.
XX New thermostable mutant B-type DNA polymerase with a Y-GG/A amino acid
XX motif between the N-terminal 3'-5' exonuclease domain and the
XX C-terminal polymerase domain in the wild type polymerase, useful for
XX synthesizing nucleic acids

```

OY 361 RYAYERNELAPNKDEEYRRRLRTTYLGYYKEBERGLWENITVLDRCPLPSITVTN 420
DB 361 rvaYernelapnkpdeeyrrrlrttylggyvkeperglwenthayldfcsjpsilvtn 420
OY 421 VSPDTLEREGCKNDVAIVGKPKDFGFPISILGELITMRQIKKKMKATIDPIEK 480
DB 421 vspdtlerEGCKNDVAIVGKPKDFGFPISILGELITMRQIKKKMKATIDPIEK 480
OY 481 MLDYKORAVFLHANSYVYMGYPKARMSKECAESVTAMGRHYIMTJKEITEKGFYVL 540
DB 481 mldYKORAVFLHANSYVYMGYPKARMSKECAESVTAMGRHYIMTJKEITEKGFYVL 540
OY 541 YADTDGYATIPGKPEPTIKKAKEFLKYINSKLPGLLEVEGFTLRGFPVAKKRYAVI 600
DB 541 yadtDgyatIpGkPePtIKKAKEfLkyINSkLPGLLeVEGfTLRGfPVAKKRYAVI 600
OY 601 DEEGITRGLLEVVRDMSLAKETQAKVLEAILKEDSVKRAVELVKNVVEIAKYQVPL 660
DB 601 deegITrGLLeVVRdMSlAKETQAKVLEAILKEDSVKRAVELVKNVVEIAKYQVPL 660
OY 661 EKLVIHEQITDLSYKAIQPHVAIAKRLAKGIVKRPSTIISYIVLRSGKISDRVILL 720
DB 661 ekLviHEqITdLSyKAIqPHVAIAKRLAKGIVKRPSTIISYIVLRSGKISDRVILL 720
OY 721 SEYDPKAKKIPDPYITENQVLPVLRILFAGYRKREDIKYSSKQVGLDAMLKK 774
DB 721 seYdpKAKKIPDPYITENQVLPVLRILFAGYRKREDIKYSSKQVGLDAMLKK 774

RESULT 3
ID AAM29322 standard; Protein; 1829 AA.
XX AAM29322;
XX
DT 20-APR-1998 (first entry)
DE DNA polymerase with 3'-5' exonuclease activity.
XX
XX TYPOI Inttron; DSM 10597; thermostable; DNA polymerase;
KM 3'-5' exonuclease; amplification.
XX
XX Thermococcus sp.
FH Key Location/Qualifiers
FT Misc-difference 1116 /label= unknown
FT /label= "encoded by GGN"
FT Misc-difference 1118 /label= unknown
FT /note= "encoded by NTC"
FT Misc-difference 1123 /label= unknown
FT /label= "encoded by NTG"
XX Del19611759-A1.
XX
XX 02-OCT-1997.
XX
XX 25-MAR-1996; 96DE-1011759.
XX
XX 25-MAR-1996; 96DE-1011759.
XX
XX (POT) MOERINGER MANNHEIM GMBH.
XX
XX Antiranikian G, Frey B, Niehaus F;
XX WPI; 1997-481494/45
XX N-PSDB; AAT86433.
XX
XX Thermostable DNA polymerase from Thermococcus sp. TY - useful for
PT nucleic acid amplification
XX

```

```

PS Claim 1; Pages 5-8; 32pp; German.
XX
XX The present sequence (TYPOI Inttron) is a Thermococcus sp. TY
CC (DSM 10597) thermostable DNA polymerase with 3'-5' exonuclease
CC activity.
CC The enzyme can specifically amplify nucleic acid fragments of up to
CC 5 kb in high yields, has an activity half-life of 20 minutes at 90
CC degrees C, has an optimum temperature of 70-80 degrees C, has an
CC optimum pH of 7.5, exhibits optimum activity at a KCl concentration
CC of 80-100 mM, is magnesium ion-dependent and is inhibited by
CC manganese ions.
XX
XX Sequence 1829 AA;
XX
Query Match 85.6%; Score 3446.5; DB 18; Length 1829;
Best Local Similarity 42.2%; Pred. No. 4,7e-260;
Matches 771; Conservative 0; Mismatches 3; Indels 1055; Gaps 3;

OY 1 MIFDFTYTKDGKPIIRIFKKEGFEKIELDPHFQPIYALTKDSDAIDEIKRGERHG 60
DB 1 mldtDytkDgKpIIRIFkKEGfEkiELDPHFqPIYAlTKdSDAIdEIKRgERHG 60
OY 61 KIVRVDAVKKKKFLGADVEMWKLIFEPHODVPALRGKIRHPAVIDIYEYDIPFAKRY 120
DB 61 kivrVdAvKKKKfLGdVEMwKLIFePHODVPALRGKIRHPAVIDIYEYDIPFAKRY 120
OY 121 LIDKGLIMEGDEELKMAFDIEFTFYHGDGFEKKEIIMISVADDEKRVITWKHIDIPY 180
DB 121 lIdKglImEGdELKMAFDIEFTFYHGDGFEKKEIIMISVADDEKRVITWKHIDIPY 180
OY 181 VDVSNEREMIKRFVOIVREKDPDLVITYNGDNFPLPYLIRAKELGVTLLGRDKHPE 240
DB 181 vdvsnERemIKrfVOIVREKdPDLVITyNGDNfPLPYLIrAKELGVTLLGRDKHPE 240
OY 241 PKIRMGDSFVETIKGRIFHFDLPVVRRTINLPYTLFAYEAVLGKTKSLGAEINAI 300
DB 241 pKIRMGdSFvETIKGRIFHFDLPVVRRTINLPYTLFAYEAVLGKTKSLGAEINAI 300
OY 241 PKIRMGDSFVETIKGRIFHFDLPVVRRTINLPYTLFAYEAVLGKTKSLGAEINAI 300
DB 241 pKIRMGdSFvETIKGRIFHFDLPVVRRTINLPYTLFAYEAVLGKTKSLGAEINAI 300
OY 301 WETESMKKLAQYSMEDARATYELGKEFPPEAEIAKLIGSVMDVSSSGNIVEWYLL 360
DB 301 wETESmKKLAQYSMEdARATyELGKEFPPEAEIAKLIGSVMDVSSSGNIVEWYLL 360
OY 361 RYAYERNELAPNKDEEYRRRLRTTYLGYYKEBERGLWENITVLDRC----- 410
DB 361 rvaYernelapnkpdeeyrrrlrttylggyvkeperglwenthayldfrcpAdkvYk 420
OY 411 ----- 410
DB 421 gKgiVnsdvKegdyIlgIdgQrvKkwkyhyegKlinIngIkctpnhkYpvtendrq 480
OY 411 ----- 410
DB 481 trIdslakSflsgkvXgkIlctklfEkIaeFknkpseEelkgelsgIlLaegtlIrK 540
OY 411 ----- 410
DB 541 dIeyfdsrYgKkrishqYrvElItgenekellertIlyfDklfgIrpSVKKgdtNaIkI 600
OY 411 ----- 410
DB 601 tTakkavYlgIeellkniesIyapavIrgfFerdatvNkIrsIvvtgtnkWiDivA 660
OY 411 ----- 410
DB 661 kIIsIgiPySrYeykyIengkelKhllEltgIdgIlIfglYgfIsseKnealeKaIe 720
OY 411 -----LYPSIIVTN 420
DB 721 vremnrlKmsfynIstfevaseYyKgevydItlegnpyfAnglltnsIypsilvtn 780
OY 421 VSPDTLEREGCKNDVAIVGKPKDFGFPISILGELITMRQIKKKMKATIDPIEK 480

```

```

Db 781 vspdtlercgcknydvaplvgykfcckdpgfipslgelitmrgeikkmkaticdplekk 840
QY 481 MLDYRQRAVKLHA----- 493
| | | | |
Db 841 mldygravavkllansllpnwplengevkfkigefidrymeeqdkrvtdntevle 900
QY 494 ----- 493
Db 901 vdnlfafslnkeeksekikvkallrhkygeaeveinsgrkhltgrhslftlrngkl 960
QY 494 ----- 493
Db 961 kelwgeevkvvgdillvppkvvklnkeavlnlpelisklpdedadvmtcpvkgtknfk 1020
QY 494 ----- 493
Db 1021 gmlrlklwlfgeeskrirtfnrylfhleelgvkllprgyevtdwglkryrqlylekvlk 1080
QY 494 ----- 493
Db 1081 nlryngnkreylvrfndikdsvscfprkeleewkixlckgfrxkcllkvdedfgkflyy 1140
QY 494 ----- 493
Db 1141 vsegyagacknktgmsysvklvnenpnvlkdmklnaektfgvrvyvgkncvdlpkkmayl 1200
QY 494 ----- 493
Db 1201 lasklcvtaenkrripalfdsesepvwaflrayfvvgdghpskrlrlstksellangl 1260
QY 494 ----- 493
Db 1261 vflnslgvasiklgfsgvyrvyinedlpflqtrgkntypnllpkveleelfgkrkfg 1320
QY 494 ----- 493
Db 1321 knltfekfeladsgkldkrkvklldfllngdivldrknvkereygyydlssvednen 1380
QY 494 ----- 493
Db 1381 flvfglllyahnsygyymgyppkarwyskceasvtawgrnyiemtlkeleekfgkflya 1440
QY 494 ----- 493
Db 1441 dsvtgdtellvkrngrieftvpleklfervdyrigekeyclledvealldnrgklwkvk 1500
QY 494 ----- 493
Db 1501 pyvmrhrakkkvyrwltnswyldvtedhslivaedgkhearpmelegkallatkdldsg 1560
QY 494 ----- 493
Db 1561 veylkphaleelisyngvydlevvegthrffangllvhnhtdgfyatlpgepeticrkake 1620
QY 494 ----- 493
Db 1621 flkylnsklpglllelegfyirgffvakkryavidegrlitrgelevrrtdseelaket 1680
QY 494 ----- 493
Db 1681 qakvleallkedsvakeveivkdveaelakypvleklvheqitkdlseykaighval 1740
QY 494 ----- 493
Db 1741 akrlaakgtvrrcttiisylvlrgskisrprvllseyopkrrkkyppdyiienovpavl 1800
QY 494 ----- 493
Db 1801 rllleaftyrtedlkqyskqvgldawlk 1829

```

RESULT 4
PAR38095

```

ID AAR38095 standard; Protein; 1702 AA.
XX
AC AAR38095;
XX
DT 21-JUN-1994 (first entry)
XX
DE T. litoralis DNA polymerase.
XX
KW DNA polymerase; intron; translation; bacteriophage NEB #619; E. coli;
KM mutation; cell growth; endonuclease; self splicing; stability;
KW expression; transformation; expression vector.
XX
OS Thermococcus litoralis.
XX
PN EP547920-A.
XX
PD 23-JUN-1993.
XX
PF 18-DEC-1992; 92EP-0311622.
XX
PR 18-DEC-1991; 91US-0811421.
XX
PA (NEMO ) NEW ENGLAND BIOLABS INC.
XX
PI Comb DG, Jack WE, Kucera R, Perlter F;
XX
DR WPI; 1993-198889/25.
XX
DR N-PSDB; AAQ43515.
XX
PT Recombinant thermostable DNA polymerase from archaeobacteria -
PT expressed efficiently by removal of intervening introns
XX
PS Claim 1; Fig 19; 63pp; English.
XX
SS
CC This sequence represents T. litoralis DNA polymerase. The DNA
CC sequence encoding this protein contains introns described in the
CC specification, however this protein sequence appears to be translated
CC through these intronic sequences. Only a portion of the DNA
CC polymerase protein sequence is reproduced in the specification. The
CC T. litoralis DNA polymerase gene sequence may be derived from
CC bacteriophage NEB #619 on an approx. 14kb BamHI restriction fragment.
CC When this gene is transformed in to E. coli and expressed in its
CC entirety, it is found to be unstable. There is a high frequency of
CC mutation in the polymerase gene, cell growth is slow and there is some
CC degree of cell mortality. This instability is due in part to the
CC presence of the two introns. It is thought that the intervening DNA
CC at this sequence encodes a separate protein which may be spliced out
CC at the protein level. The second, 1170bp, intron has been found to
CC code for an endonuclease which self splices out in E. coli. To
CC improve stability of the DNA when it is expressed in E. coli, the
CC specification states that it is desirable to delete the introns
CC before transformation of the expression vector into E. coli.
XX
SQ Sequence 1702 AA;

```

Query Match 83.4%; Score 3356; DB 14; Length 1702;
Best Local Similarity 42.7%; Pred. No. 5e-253;
Matches 726; Conservative 33; Mismatches 15; Indels 928; Gaps 2;

```

QY 1 MFEDDYIRKDKSPIIRIRKKEGFEKIEIDPPOPIYALKKDSADIEIKRIGERHG 60
| | | | |
Db 1 mldtdyikdgkplirirkkegefkelophtgpyllyalkdsaleelalgerng 60
| | | | |
QY 61 KIVRVDAVAVKVKKELGRDVEVWKLFEHPDPVALRGKIRHPAVIDIEYDIPFARKY 120
| | | | |
Db 61 kvrvldavkvkklfgrvevwmklifehpqdvpmrgklrehpavldleydipfarky 120
| | | | |
QY 121 LLDKGLIPMEGDEELKNAFDIETFYHSGDERGKGIIMISYADDEBARVITWKINDLPY 180
| | | | |
Db 121 lldkgllpmegdeellkilaftietfyhsgdefgkgeilmisyadeearvltwnkndipy 180
| | | | |
QY 181 VDVVSNREMIKRFQIVREKDPDLITVNGNPNFLPLIKRAEKLGVTLIGRKEHPE 240
| | | | |

```

Db	1261	lkqklllepikcyvsnnyyknegodfnllakslv	fmkhrkthoekgrkkipefmyelpv	1320
Qy	544	-----	-----	543
Db	1321	tyleaffjrglfsadgvtvlrkyvpeirltnldadfi	revrkllwlvgsnslfaetpcnr	1380
Qy	544	-----	-----	543
Db	1381	yngvstgltyskhlrlknkwrfaerlgflerqkrl	lehlksarvkrntldfgldlvhwk	1440
Qy	544	-----	-----	572
Db	1441	kveeipjegyydiaveethrffannllvhncdg	fiactlpgskpelikkakeflnyns	1500
Qy	573	klpglelleeyrbegylkgsfynakrrvavidecgr	itttgsleevvrndmseiatetoakvlea	632
Db	1501	klpglelleeyrgfifrlckrryavideegflltr	lgjevrrcrwseialetqakvlea	1560
Qy	633	llkdsyvekaneylkdvveeiarayovrpleklyv	hbedtmkdssevkatoiphuavikrrlaak	692
Db	1561	llkgsyvekaevvrdvveiaikyavpleklyv	ltheqiltldldykajagphvialakrlaar	1620
Qy	693	gikvrpgtllisylrncsgklsidrvlllseypdk	kxhkydpdyiyinoyulpavlrlilaafg	752
Db	1621	gikvrpgtllisylrncsgklsidrvlllseypd	krkhkydpdyiengvlpavrlilaafg	1680
Qy	753	yrkedlkrtyssokvgvndamluk	774	
Db	1681	yrkedlkrtyssokvgvndamluk	1702	

RESULT	
ID	AAMW6285
AC	AAMW6285 standard; Protein; 773 AA.
DJ	AAMW6285;
EI	28-SEP-1998 (first entry)
FH	Thermococcus gorgonarius (Tgo) DNA polymerase protein.
GK	Thermostable: DNA polymerase; Thermococcus gorgonarius; Tgo; screening; 3'-5' proofreading exonuclease; Taq polymerase; PCR; amplification; intervening sequence.
HU	Thermococcus gorgonarius.
IQ	EPB34570-A1.
JL	08-APR-1998.
KM	03-OCT-1996; 96EP-0115874.
LN	03-OCT-1996; 96EP-0115874.
MV	(BOEF) BOEHRINGER MANNHEIM GMBH.
NX	Angeker B., Ankenbauer W., Bonch-osmojovskaya E., Ebenichler C.; Laue F., Schmitz-agheguyan G., Svetlichny V. MPI: 1998-195468/18, N-PDSB: AAV26371.
OY	New isolated thermostable DNA polymerase - obtained from Thermococcus gorgonarius, used for amplifying DNA or for DNA cloning, sequencing or labeling
PZ	Disclosure: Page 16-21; 41pp: English.
R1	This amino acid sequence produces the thermostable DNA polymerase in Thermococcus gorgonarius (Tgo). The DNA polymerase enzyme was isolated from this species by standard isolation and purification techniques. The thermostable enzyme possesses a 3'-5' proofreading exonuclease function.

CC this activity is not present in Tag polymerase which is normally used in
 CC PCR amplification. Thus this activity enables the fidelity of PCR to be
 CC increased, and therefore would no longer be prone to base incorporation.
 CC The Tgo DNA polymerase exhibits a two fold greater replication fidelity
 CC than any known polymerases. It can also decrease non-specific background
 CC amplification in PCR, due to its 3'-5' activity. The gene from this
 CC species also does not contain intervening sequences, which would have to
 CC be removed to enable expression in *E. coli* to occur. This enzyme has an
 CC approx molecular weight of 92-96 kD, and has been shown to retain 90% of
 CC its activity after incubation, in the presence of a stabilizer for two
 CC hours at 95 deg. C. This Tgo DNA polymerase can be used for e.g.
 CC amplifying DNA or for DNA cloning, sequencing or labelling.

XX Sequence 773 AA:

Query Match 80.3%; Score 3232.5; DB 19; Length 773;
 Best Local Similarity 77.7%; Pred. No. 7.2e-244;
 Matches 601; Conservative 83; Mismatches 86; Indels 3; Gaps 2;

```

QY 1 MFEDDYTKDKGPIIRIKKENGGEFKIELDPHFOPIYALLKDSADIEKAIGERHG 60
DB 1 mldtdytedgkpvrltkkenggeftldydrnfepylallkdsdpdvkkaetnng 60
QY 61 KIVRVDAVKVKKKFLGRDVEWVKLFEHPDVPALRGKIREHPAVIDIEYDIPAKRY 120
DB 61 tlvrvvraekvkkkflgrplevwklyftbpgdvpatrikhehpavvdlyeydipakry 120
QY 121 LIDKGLIPMEGDEELKLMFDIETFYHSGDEFGKEIIMISYADEBARVITWKNIDLPY 180
DB 121 lldkglipmegdeeklmfadietlyhegeefaeagplimisyadeagavltwknidpy 180
QY 181 VDVSNEREMIKRFVOIYREKDPDLITVNGDNFDPYLIRAKELGVTLLGRDKEHPE 240
DB 181 vdvsnereimikrfvoiyrekdpdlitvngdnfpylirakelgvttllgrdkehpe 240
QY 241 PPIHMGDSFAVEIKRIFHEDELFPVVRRTINLPTYLLEAVIRGAVIKTSKIGAEIRAI 300
DB 241 ppihmgdsfaveikrifehdelpvvrirtinlptylleavirgaviktskigaeirai 300
QY 239 PPIGTMGDFAVEVGRITFDIDYVTRITINLPTYLEVYAEIIGPQKAEYAEIQA 298
DB 239 ppigtmgdfavevgritfdidypvtritinlptylevyaeeiigpqkaeyaeiqa 298
QY 301 WETESMKRLAOYSMEDARATYELGKEFPMEAEIAGISQSWDVSRSSTGNLVEWYLL 360
DB 301 wetesmkrlaoysmedaratyelgkefpmeaeiagisqswdvsrsstgnlvewyl 360
QY 299 wetgglelvarysmedavelyelgkeffmeaqslrlyvgsltwdsrstgnlwevll 358
DB 299 wetgglelvarysmedavelyelgkeffmeaqslrlyvgsltwdsrstgnlwevll 358
QY 361 RYAVENELAPKPKDEEERRLRTTYLGIVYKEPERGMENTIYDFECLIPSTIVTN 420
DB 361 ryavenelapkpdeeerlrtrtylgivykepergmentiydfeclypstivtn 420
QY 359 rkayernelapkpderelarf-resyaggykpergilwenlyldtrslpsliltn 417
DB 359 rkayernelapkpderelarf-resyaggykpergilwenlyldtrslpsliltn 417
QY 421 VSPDTEREGCKNVDYAPIVGKFCDFPGFIPSIIGELITRHOETKKMKATIDPIEKK 480
DB 421 vspdteregcknvdypivgkfcdfpgfipslilgdlleerqvkmmkatiidpiekk 480
QY 418 vspdtlnegeceydvapgvnkhfckdipglipslilgdlleerqvkmmkatiidpiekk 477
DB 418 vspdtlnegeceydvapgvnkhfckdipglipslilgdlleerqvkmmkatiidpiekk 477
QY 481 MLDYRQAVAKLHANSYYGYMKPKARWSEKCAESVTANGRYHIENTIKIEEKEGFKVL 540
DB 481 mldyrqavaklhansyygytkarwykceasvtagryhientikieekegfkvl 540
QY 478 lldygralkilansyygytkarwykceasvtagryhientikieekegfkvl 537
DB 478 lldygralkilansyygytkarwykceasvtagryhientikieekegfkvl 537
QY 541 YADTGSFATITGEKPEITKKAKERFLKYNLSKPGLELEHFGFLRKFFPAKKRYAVI 600
DB 541 yadtgsfatitgeketitkkakerflkynlskpglelehgflrkffpaakryavi 600
QY 538 yadtdgfatitlpgdaetkkekafldylnaklpallelegyfkrgfvtlckkyavi 597
DB 538 yadtdgfatitlpgdaetkkekafldylnaklpallelegyfkrgfvtlckkyavi 597
QY 601 DEEGRTTGTGLLEVRDMEIARETOAKVLEALILKEDSVSEIYKDVDEEIAKQVQL 660
DB 601 deegrttgtgllevrddmeiaretoakvlealilkedsvsei ykdvdeeeiakqvql 660
QY 598 deegrttgtgllevrddmeiaretoakvlealilkedsvsei ykdvdeeeiakqvql 657
DB 598 deegrttgtgllevrddmeiaretoakvlealilkedsvsei ykdvdeeeiakqvql 657
QY 661 EKLVIHQITKDLSEYKAIGPVAIALAKRLAAGIKVRCPTIISYVLRSGKISDVILL 720
DB 661 eklviahqitkdlseykaigpvaialakrlaagikvrcptiisylvlrsgkisdvill 720
QY 658 eklviyegtrtdlkdykagphvavakriaargiklrptvlsyvlilksqslgdratf 717
DB 658 eklviyegtrtdlkdykagphvavakriaargiklrptvlsyvlilksqslgdratf 717
QY 721 SEYDPKHHYDDPYITENOVLPVAILERFAGYRKEDLKYQSSKQVGLDAMLK 773
DB 721 seydpkhhyyddpyitenovlpvailerfagyrkedlkyqsskqvgladamlk 773
QY 718 defdpakhkydeeylengvlpaverlirafigyrkedlkyqtrgyglgawlk 770
DB 718 defdpakhkydeeylengvlpaverlirafigyrkedlkyqtrgyglgawlk 770

```

RESULT 6

AAW58818

ID AAW58818 standard; Protein; 773 AA.

XX AAW58818;

XX 17-AUG-1998 (first entry)

XX Thermococcus gorgonarius thermostable DNA polymerase.

XX DNA polymerase; thermostable enzyme; DNA amplification; PCR;

XX cloning; sequencing; DNA labelling.

XX Thermococcus gorgonarius.

XX EP834571-A1.

XX 08-APR-1998.

XX 16-JAN-1997; 97EP-0100584.

XX 03-OCT-1996; 96EP-0115874.

XX (BOE) BOEHRINGER MANNHEIM GMBH.

XX Angerer B, Ankenbauer W, Bonch-Osmolovskaya E, Ebenbichle RC;

XX Lane F, Schmitz-Agnehan G, Svetlichny V;

XX WPI: 1998-195469/18.

XX N-PSDB: AAV11396.

XX New isolated thermostable DNA polymerase - obtained from

XX Thermococcus gorgonarius, used for amplifying DNA or for DNA

XX cloning, sequencing or labelling.

XX Claim 1; Fig 3; 41pp: English.

CC This polypeptide comprises a novel thermostable DNA polymerase
 CC (TDP) of *Thermococcus gorgonarius*, a thermophilic archaeobacterium
 CC isolated from a thermal vent. The TDP catalyses the directed
 CC polymerisation of DNA and retains about 90% of its activity after
 CC incubation for 2 hr at about 95 deg C in the presence of a
 CC stabilizer. It exhibits more than 2-fold greater replication
 CC fidelity than DNA polymerase from *Pyrococcus furiosus*, and also has
 CC 3'-5' exonuclease (proofreading) activity. This exonuclease
 CC activity can decrease non-specific background amplification in PCR
 CC by degrading delayed ends of primers bound to unspecific
 CC sequences. The TDP is suitable for use in DNA amplification,
 CC cloning, sequencing and labelling. The native enzyme can be used,
 CC or a recombinant enzyme that has been expressed in microbial
 CC host cells transformed with a vector carrying an isolated DNA
 CC sequence (see AAV11396) coding for TDP.

XX Sequence 773 AA:

Query Match 80.3%; Score 3231.5; DB 19; Length 773;
 Best Local Similarity 77.7%; Pred. No. 8.6e-244;
 Matches 601; Conservative 83; Mismatches 86; Indels 3; Gaps 2;

```

QY 1 MFEDDYTKDKGPIIRIKKENGGEFKIELDPHFOPIYALLKDSADIEKAIGERHG 60
DB 1 mldtdytedgkpvrltkkenggeftldydrnfepylallkdsdpdvkkaetnng 60
QY 61 KIVRVDAVKVKKKFLGRDVEWVKLFEHPDVPALRGKIREHPAVIDIEYDIPAKRY 120
DB 61 tlvrvvraekvkkkflgrplevwklyftbpgdvpatrikhehpavvdlyeydipakry 120
QY 121 LIDKGLIPMEGDEELKLMFDIETFYHSGDEFGKEIIMISYADEBARVITWKNIDLPY 180
DB 121 lldkglipmegdeeklmfadietlyhegeefaeagplimisyadeagavltwknidpy 180
QY 181 VDVSNEREMIKRFVOIYREKDPDLITVNGDNFDPYLIRAKELGVTLLGRDKEHPE 240
DB 181 vdvsnereimikrfvoiyrekdpdlitvngdnfpylirakelgvttllgrdkehpe 240

```



```

Db      181 |dvvnst|em|k|rf|k|v|k|k|d|p|v|l|l|y|n|g|n|d|f|a|y|l|k|k|r|e|k|g|y|k|f|l|g|r|---|e|g|s|e| 238
Oy      241 |P|H|M|G|S|F|A|E|I|G|R|H|F|D|F|P|V|R|R|T|I|N|P|T|Y|L|E|A|V|Y|A|V|G|K|R|K|S|K|A|G|E|I|A|I| 300
Db      239 |p|l|q|m|g|r|f|a|v|e|g|r|h|f|d|l|y|r|l|r|l|n|p|y|l|e|a|v|e|a|l|g|q|k|e|k|y|a|e|e|l|a| 298
Oy      301 |W|E|T|E|S|M|K|L|A|Q|Y|S|M|E|D|R|A|T|Y|E|L|G|E|F|F|P|E|A|E|L|A|K|I|G|O|S|V|M|D|V|S|S|T|G|N|L|E|W|Y|L| 360
Db      299 |w|e|t|s|e|g|l|e|r|v|a|r|y|s|m|e|d|k|v|y|e|l|g|e|f|f|m|e|a|q|s|r|l|v|g|s|l|w|d|v|s|r|s|s|t|g|n|l|e|w|f|l| 358
Oy      361 |R|V|A|R|R|N|E|L|A|P|N|K|P|D|E|E|Y|R|R|R|L|R|T|Y|L|G|Y|V|K|E|P|R|G|M|E|N|T|Y|L|D|F|R|C|L|Y|S|I|I|V|H|N| 420
Db      359 |r|k|a|y|e|r|n|e|l|a|p|n|k|p|d|e|r|a|r|-r|e|s|y|e|g|y|k|e|p|e|r|g|w|e|n|i|v|y|l|d|f|r|s|l|y|s|i|l|t|h|n| 417
Oy      421 |V|S|P|D|L|R|E|G|C|K|N|Y|A|P|R|V|G|K|F|C|K|D|P|G|R|F|P|S|I|L|G|E|I|T|M|R|O|R|E|I|K|K|R|K|M|A|T|D|I|R|E|K| 480
Db      418 |v|s|p|d|l|r|e|g|c|e|e|y|v|a|p|r|v|g|n|k|f|c|k|d|p|g|r|f|p|s|i|l|g|e|l|e|e|r|g|k|r|k|m|k|a|t|d|i|p|l|e|k| 477
Oy      481 |M|L|D|Y|R|O|R|A|V|K|L|H|A|N|S|Y|G|Y|G|P|K|R|A|W|Y|S|K|E|C|A|E|S|V|T|A|G|R|H|Y|I|M|T|K|E|I|E|K|F|G|F|V|L| 540
Db      478 |l|l|d|y|r|o|r|a|v|k|l|h|a|n|s|f|y|g|y|g|t|k|a|r|w|y|y|k|e|a|e|s|v|t|a|g|r|e|y|i|e|t|l|r|e|l|e|e|k|f|g|f|v|l| 537
Oy      541 |Y|A|D|T|G|F|Y|A|T|P|G|E|P|E|T|I|K|K|A|E|F|L|K|Y|I|N|S|K|L|P|G|L|E|L|E|Y|E|G|F|Y|L|G|F|V|A|K|R|R|A|V|I| 600
Db      538 |y|a|d|t|g|f|a|t|p|g|a|d|e|t|v|k|k|k|a|e|f|l|y|i|n|a|k|l|p|g|l|e|l|e|y|e|g|f|y|k|r|g|f|v|t|k|k|k|y|a|v|i| 597
Oy      601 |D|E|E|G|R|I|T|R|G|L|E|V|R|D|M|S|E|I|A|K|E|T|Q|A|K|V|L|E|A|I|L|K|E|D|S|V|E|R|A|V|E|L|K|Y|V|E|I|A|Y|Q|V|P|L| 660
Db      598 |d|e|e|g|r|i|t|r|g|l|e|v|r|d|m|s|e|i|a|k|e|t|q|a|v|l|e|a|i|l|k|e|d|s|v|e|r|a|v|e|l|k|y|v|e|i|a|y|q|v|p|l| 657
Oy      661 |E|K|V|I|H|E|O|I|R|D|L|S|E|Y|K|A|I|G|P|H|V|A|I|A|K|R|L|A|K|R|A|G|R|K|R|G|T|I|S|Y|I|V|L|G|S|G|R|S|R|V|I|L| 720
Db      658 |e|k|v|i|h|e|o|i|r|d|l|s|e|y|k|a|i|g|p|h|v|a|i|a|k|r|l|a|k|r|a|g|r|k|r|g|t|i|s|y|i|v|l|g|s|g|r|s|r|v|i|l| 717
Oy      721 |S|E|Y|D|P|R|K|H|K|Y|D|P|D|Y|I|E|N|O|V|L|P|A|V|R|I|L|E|A|F|G|R|K|E|D|I|K|Y|O|S|S|K|O|Y|G|D|A|W|L|K| 773
Db      718 |d|e|f|d|p|k|h|k|y|d|a|e|y|i|e|n|v|l|p|a|v|r|i|l|r|a|f|g|r|k|e|d|i|r|y|k|t|r|g|v|g|i|s|a|w|l|k| 770

RESULT 7
AAW33111
ID      AAW33111 standard; Protein: 774 AA.
XX
XX      AAW33111:
XX
XX      28-JAN-1998 (first entry)
XX
XX      Thermococcus peptonophilus OG-1 thermostable DNA polymerase.
XX
XX      JCM 9653; thermostable; DNA polymerase;
XX      nucleic acid sequence amplification; PCR.
XX
XX      Thermococcus peptonophilus.
XX
XX      JF09252776-A.
XX
XX      30-SEP-1997.
XX
XX      19-MAR-1996; 96JP-0063112.
XX
XX      19-MAR-1996; 96JP-0063112.
XX
XX      19-MAR-1996; 96JP-0063112.
XX
XX      (TOYM ) TOYOBO KK.
XX
XX      WPI: 1997-530149/49.
XX
XX      N-PSDB; AAT88373.
XX
XX      Thermococcus peptonophilus thermostable DNA polymerase - useful for
XX      PT      nucleic acid sequence amplification, e.g. polymerase chain reaction
XX
XX      Claim 9; Pages 9-12; 27pp; Japanese.
XX

```

```

CC      The present sequence is a Thermococcus peptonophilus OG-1
CC      (JCM 9653) derived thermostable DNA polymerase, which can be used
CC      for nucleic acid sequence amplification, e.g. PCR.
XX
XX      SQ      Sequence 774 AA:

Query Match      80.1%; Score 3226.5; DB 18; Length 774;
Best Local Similarity 77.7%; Pred. No. 2.1e-243;
Matches 601; Conservative 82; Mismatches 87; Indels 3; Gaps 2;

Oy      1 |M|I|F|D|Y|T|T|K|D|K|P|R|I|T|I|F|K|K|E|G|E|F|I|E|D|P|F|O|P|R|Y|A|L|K|D|S|A|I|D|E|I|A|K|I|G|R|E|H|G| 60
Db      1 |m|i|l|d|t|d|y|t|e|n|g|k|p|r|i|f|i|k|k|e|n|g|e|f|k|e|y|d|t|f|e|p|y|l|a|l|k|d|s|a|l|e|e|v|k|l|a|e|h|g| 60
Oy      61 |K|I|V|R|V|D|A|V|K|K|K|F|L|G|R|D|V|E|W|K|L|F|E|H|P|O|V|P|R|L|R|G|R|K|R|H|P|A|V|I|D|I|E|Y|D|I|P|A|R|K|Y| 120
Db      61 |k|i|v|r|v|d|a|v|k|k|k|f|l|g|r|d|v|e|w|k|l|f|e|h|p|o|v|p|r|l|r|g|r|k|r|h|p|a|v|i|d|i|e|y|d|i|p|a|r|k|y| 120
Oy      121 |L|I|D|K|L|I|M|E|G|D|E|L|K|L|A|F|D|I|E|T|F|Y|H|G|D|E|R|G|K|E|I|M|I|S|V|A|D|E|E|R|A|R|V|T|W|K|R|I|D|L|P|Y| 180
Db      121 |l|i|d|k|l|i|m|e|g|d|e|l|k|l|a|f|d|i|e|t|f|y|h|g|d|e|r|g|k|e|i|m|i|s|v|a|d|e|e|r|a|r|v|t|w|k|r|i|d|l|p|y| 180
Oy      181 |V|D|V|S|N|E|R|E|M|T|K|R|F|V|Q|I|V|R|E|K|D|P|V|L|I|T|Y|N|G|D|N|F|D|P|Y|L|I|K|R|A|E|K|I|G|Y|T|L|L|R|D|R|K|H|P|E| 240
Db      181 |v|d|v|s|n|e|r|e|m|t|k|r|f|v|q|i|v|r|e|k|d|p|v|l|i|t|y|n|g|d|n|f|d|p|y|l|i|k|r|a|e|k|i|g|y|t|l|l|r|d|r|k|h|p|e| 238
Oy      241 |P|H|M|G|S|F|A|E|I|G|R|H|F|D|F|P|V|R|R|T|I|N|P|T|Y|L|E|A|V|Y|A|V|G|K|R|K|S|K|A|G|E|I|A|I| 300
Db      239 |p|h|g|m|g|s|f|a|e|i|g|r|h|f|d|f|p|v|r|r|t|i|n|p|t|y|l|e|a|v|y|a|v|g|k|r|k|s|k|a|g|e|i|a|i| 300
Oy      301 |W|E|T|E|S|M|K|L|A|Q|Y|S|M|E|D|R|A|T|Y|E|L|G|E|F|F|P|E|A|E|L|A|K|I|G|O|S|V|M|D|V|S|S|T|G|N|L|E|W|Y|L| 360
Db      299 |w|e|t|s|e|g|l|e|r|v|a|r|y|s|m|e|d|k|v|y|e|l|g|e|f|f|m|e|a|q|s|r|l|v|g|s|l|w|d|v|s|r|s|s|t|g|n|l|e|w|f|l| 358
Oy      361 |R|V|A|R|R|N|E|L|A|P|N|K|P|D|E|E|Y|R|R|R|L|R|T|Y|L|G|Y|V|K|E|P|R|G|M|E|N|T|Y|L|D|F|R|C|L|Y|S|I|I|V|H|N| 420
Db      359 |r|k|a|y|e|r|n|e|l|a|p|n|k|p|d|e|r|a|r|-r|e|s|y|e|g|y|k|e|p|e|r|g|w|e|n|i|v|y|l|d|f|r|s|l|y|s|i|l|t|h|n| 417
Oy      421 |V|S|P|D|L|R|E|G|C|K|N|Y|A|P|R|V|G|K|F|C|K|D|P|G|R|F|P|S|I|L|G|E|I|T|M|R|O|R|E|I|K|K|R|K|M|A|T|D|I|R|E|K| 480
Db      418 |v|s|p|d|l|r|e|g|c|e|y|d|t|a|p|r|v|g|n|r|f|c|k|d|p|g|r|f|p|s|i|l|g|e|l|e|e|r|g|k|r|k|m|k|a|t|d|i|p|l|e|k| 477
Oy      481 |M|L|D|Y|R|O|R|A|V|K|L|H|A|N|S|Y|G|Y|G|P|K|R|A|W|Y|S|K|E|C|A|E|S|V|T|A|G|R|H|Y|I|M|T|K|E|I|E|K|F|G|F|V|L| 540
Db      478 |l|l|d|y|r|o|r|a|v|k|l|h|a|n|s|f|y|g|y|g|t|k|a|r|w|y|y|k|e|a|e|s|v|t|a|g|r|e|y|i|e|t|l|r|e|l|e|e|k|f|g|f|v|l| 537
Oy      541 |Y|A|D|T|G|F|Y|A|T|P|G|E|P|E|T|I|K|K|A|E|F|L|K|Y|I|N|S|K|L|P|G|L|E|L|E|Y|E|G|F|Y|L|G|F|V|A|K|R|R|A|V|I| 600
Db      538 |y|a|d|t|g|f|a|t|p|g|a|d|e|t|v|k|k|k|a|e|f|l|y|i|n|a|k|l|p|g|l|e|l|e|y|e|g|f|y|k|r|g|f|v|t|k|k|k|y|a|v|i| 597
Oy      601 |D|E|E|G|R|I|T|R|G|L|E|V|R|D|M|S|E|I|A|K|E|T|Q|A|K|V|L|E|A|I|L|K|E|D|S|V|E|R|A|V|E|L|K|Y|V|E|I|A|Y|Q|V|P|L| 660
Db      598 |d|e|e|g|r|i|t|r|g|l|e|v|r|d|m|s|e|i|a|k|e|t|q|a|v|l|e|a|i|l|k|e|d|s|v|e|r|a|v|e|l|k|y|v|e|i|a|y|q|v|p|l| 657
Oy      661 |E|K|V|I|H|E|O|I|R|D|L|S|E|Y|K|A|I|G|P|H|V|A|I|A|K|R|L|A|K|R|A|G|R|K|R|G|T|I|S|Y|I|V|L|G|S|G|R|S|R|V|I|L| 720
Db      658 |e|k|v|i|h|e|o|i|r|d|l|s|e|y|k|a|i|g|p|h|v|a|i|a|k|r|l|a|k|r|a|g|r|k|r|g|t|i|s|y|i|v|l|g|s|g|r|s|r|v|i|l| 717
Oy      721 |S|E|Y|D|P|R|K|H|K|Y|D|P|D|Y|I|E|N|O|V|L|P|A|V|R|I|L|E|A|F|G|R|K|E|D|I|K|Y|O|S|S|K|O|Y|G|D|A|W|L|K| 773
Db      718 |d|e|f|d|p|k|h|k|y|d|a|e|y|i|e|n|v|l|p|a|v|r|i|l|r|a|f|g|r|k|e|d|i|r|y|k|t|r|g|v|g|i|s|a|w|l|k| 770

RESULT 8
AAW41313
ID      AAW41313 standard; Protein: 774 AA.
XX
XX      AAW41313:
XX
XX      20-MAY-1998 (first entry)
XX
XX      Heat-resistant DNA polymerase.
XX

```

XX Heat-resistant DNA polymerase; 3'-5' exonuclease activity;
 KW nucleic acid amplification.
 XX
 OS Pyrococcus sp.
 FH Key Location/Qualifiers
 FT Misc-difference 140..144
 FT /note="motif mutated to give polymerase of the
 FT Invention"
 XX
 PN JPI0042871-A.
 PD 17-FEB-1998.
 XX
 PF 29-JUL-1996; 96JP-0198910.
 XX
 PR 29-JUL-1996; 96JP-0198910.
 XX
 PA (TOYO) TOYOBO KK.
 XX
 DR WPI: 1998-186867/17.
 DR N-PSDB: AAV14524.
 XX
 PT New heat-resistant DNA polymerase - useful for, e.g. increasing
 PT efficiency of nucleic acid amplification
 XX
 PS Claim 3: Page 17-19; 22pp; Japanese.
 XX
 CC This sequence represents the wild-type heat-resistant
 CC DNA polymerase of the invention (HRDP), which has a 3'-5' exonuclease
 CC activity. In the HRDP of the invention, the motif shown in AAM41315 is
 CC mutated. The enzyme can increase the amplification efficiency, of
 CC nucleic acid amplification methods.
 XX
 SO Sequence 774 AA:

Query Match 80.1%; Score 3225.5; DB 19; Length 774;
 Best Local Similarity 77.5%; Pred. No. 2,5e-243;
 Matches 599; Conservative 82; Mismatches 89; Indels 3; Gaps 2;

QY 1 MFEDPTIRKDKGRIIRIKKENGKEKELDPHFOPRYIALKKDSADIEIKAKIGERRG 60
 DB 1 mldtlytedgkprvrlrtfkkengetkleydrfeyfyallddsaleevkkitaerf 60
 QY 61 KIVRVVDAAVKKKKFLGRDVEWKLIFEHPODYALRGKIREHPAVIDIYEDIPFAKRY 120
 DB 61 tvvtvkrvekvqkflgrvewwkllyftnpgdparlrdkirehgvldlyeydlfakry 120
 QY 121 LIDKGLIPHEGDEELKLMARDIETFYHEDGDEFGKELIMISVADDEARVITMKNIDLPY 180
 DB 121 lldkglypmegdeelmldfietllyhegeefaeagpilmisyadeagavltlknvdlpy 180
 QY 181 VDVVSNEREMIKRFVOIVEKDPDLITVNGDNFDLPYLIRAKKIGVTLLEGROKHP 240
 DB 181 vdvvsneremikrfvovivekdpdlitvngdnfdlpylirakrkgvttllegrokhpe 240
 QY 241 PKIHRGDSFAVEIKIRIFEDLPVVRRTINLTFTYTLAEVAENVALGTSKLAEGRIAI 300
 DB 241 pkirgdsfaveikirifedlpvvrtrintlfttytlaevaevalgtsklaegriai 300
 QY 239 pkqimgdrfavevgrhfdlpyvrlrtlnlptlyleavfavfgqpkexvaeeltpa 298
 DB 239 pkqimgdrfavevgrhfdlpyvrlrtlnlptlyleavfavfgqpkexvaeeltpa 298
 QY 301 WETESMKLAOYSMEDARATYELGKEFPFMEALAKLIGOSWDVSRSTGMLVWYLL 360
 DB 301 wetesmklaoysemedaratyelgkefpfmealaklignoswdvsrstgmlvwyl 360
 QY 361 RVAYENELAPNKPDEEVRRRRTYIGGYVKEPERGKLMENTIYDPRCLYSIIIVTH 420
 DB 361 rvayenelapnkpdeevrrrttyiggyvkepergklmentiydprclysiiivth 420
 QY 421 VSPDTLEREGCKNVDAPYVYKFCDFGPIPSILGELITMROETIKKKKATIDPIEK 480
 DB 421 vspdtlercgcknvdapyvykfcdfgpiipsilgelitmroetikkkkatidpiek 480
 QY 418 vepdclnrgcgckeydvapvyghrfckdfpgflpsllgdlleerqklkkkakatidpiek 477

QY 481 MDYRORAVKLANSYGYGYPKARWSECAESVTANGRHYTEINTEIEKFGKVL 540
 DB 478 lldyrgalrkllansyygyygararwykeceesvltangrhytintlkleekyfkvl 537
 QY 541 YADTGGFYATTIGEEKPRTTKKAKKEFLKYNKSLPGLLEFEGFLRPFPAKKRYANI 600
 DB 538 ysdtdgfialtpgdaetvkkameflnylnklpaleeyegfkyrgfylvkkkyavl 597
 QY 601 DEGRITTRGLEVRDRDWEIARETOAKVLEALIKDSVEKAVEIKDVVEETAKYQVL 660
 DB 598 deegrittrglevdrdweiaretoakvlealikdsvekaveikdvveetakyqvp 657
 QY 661 EKLVIHEQITKDLSEYKALGPVAVIAKRLANGIKYRPGTIISYVLKSGSKISDRVLL 720
 DB 658 eklvheqitrdlkdylkacgphavakrlaargvklrpgtvisylvlksgrigdrafp 717
 QY 721 SEYDPKKHKYDDPYITENOVLPVAVLRIEAPFGKRDKLKQSSKQGLDAMLK 773
 DB 718 defdpkhhkyddepyitenvlpvavlrileapfgkrdklkqsskqglamawk 770

RESULT 9
 AAR93154
 ID AAR93154 standard; Protein; 774 AA.
 AC AAR93154;
 XX
 DT 20-JUN-1996 (first entry)
 XX
 DE Mature DNA polymerase from superthermophilic archaeobacterium KOD1.
 XX
 KW DNA polymerase; superthermophilic archaeobacterium; amplification; primer;
 KW PCR; Pyrococcus furiosus; probe; E.coli; expression vector;
 XX thermal stability.
 XX
 OS Synthetic.
 XX
 PN JP07298879-A.
 XX
 PD 14-NOV-1995.
 XX
 PR 09-MAY-1994; 94JP-0095109.
 XX
 PA (TOYO) TOYOBO KK.
 XX
 DR WPI: 1996-026421/03.
 XX
 PT DNA polymerase from a super-thermophilic archaeobacterium - is
 PT suitable for polymerase chain reaction
 XX
 PS Claim 4: Page 12-14; 20pp; Japanese.
 XX

This is the amino acid sequence of the mature DNA polymerase from the
 superthermophilic archaeobacterium KOD1. The protein was constructed
 from the complete gene (AAT10258) by removing the 2 intervening
 sequences (nucleotides 1374-2453 and 2708-4316) by fusion PCR using the
 primers AAT10262-7. The resultant fragment was cloned into the
 expression vector pET-8c to produce pET-pol for expression and
 purification of the protein in E.coli.
 CC The complete gene was isolated by amplifying a fragment (AAT10261) using
 CC primers AAT10259-60 which were based on conserved amino acid sequence
 CC from the Pyrococcus furiosus DNA polymerase. The amplified fragment was
 CC used as a probe to isolate the full length gene from a Southern
 CC hybridisation on chromosomal DNA. The gene was cloned into plasmid pBS
 CC and used to transform E.coli JM109 from which a plasmid contg. the full
 CC length gene was isolated. Cells contg. the recombinant expression
 CC plasmid were used to express and purify the KOD1 DNA polymerase which is
 CC suitable for PCR due to its high thermal stability.
 XX
 SO Sequence 774 AA:

Query Match	80.0%;	Score 3222.5;	DB 17;	Length 774;
Best Local Similarity	77.4%;	Pred. No. 4.4e-243;		
Matches 598;	Conservative 83;	Mismatches 89;	Indels 3;	Gaps 2;

[illegible]

OS Hyperthermophilic archaeon strain KOD1.
 PN EP745675-A2.
 PD 04-DEC-1996.
 XX
 PF 30-MAY-1996; 96EP-0108613.
 XX
 PR 31-MAY-1995; 95JP-0134096.
 XX
 PA (TOYM) TOYO BOSEKI KK.
 XX (TOYM) TOYOBO KK.
 PI Arakawa T, Imanaka T, Inoue H, Kawakami B, Kawamura Y;
 PI Kitabayashi M, Morikawa M, Takagi M;
 XX
 DR WPI; 1997-013698/02.
 XX
 DR N-PSDB; AAT71296.
 XX
 PT New isolated thermostable KOD polymerase - from hyper:thermophilic
 XX archaeon strain KOD1, used for amplifying target nucleic acids
 PS
 PS Claim 3; Page 12-15; 44pp; English.
 XX
 CC AAW20049 is a thermostable DNA polymerase (TDP) derived from the
 CC hyperthermophilic archaeon strain KOD1. The DNA sequence encoding this
 CC enzyme contains intervening sequences, which are removed to leave a
 CC sequence encoding the polymerase. This sequence is a truncated version
 CC of the entire product of the gene (see AAW20048). The TDP has a DNA
 CC extension rate of at least 30 bases per second and a 3'-5' exonuclease
 CC activity. The TDP is used for amplifying target nucleic acid with a
 CC short reaction time and high fidelity.
 XX
 SQ Sequence 774 AA:

Query Match	80.08;	Score 3222.5;	DB 18;	Length 774;
Best Local Similarity	77.48;	Pred. No. 4.4e-243;		
Matches 598; Conservative	83;	Mismatches 89;	Indels 3;	Gaps 2;

QY	1	MLPBDYITTKOSKPLIRIRIFKKNNGFKIELDRPHORPYITALLKODSALDEIKAIKGENHG	60
Db	1	mlldcytledqrvyrirfkkengefkleydctfeyrallyalkkdsaleewkxilaethg	60
QY	61	KIVRPVADWAKVKKRFLGDNVEYWKILFEHPDOVPALRCKIREHPNVIDIYENDIPFARKY	120
Db	61	lvtyckkrxekvgkklfkgprvewklyfcthpqdvprairkctehngavldiyeydiprfarky	120
QY	121	LIDKGLPMEGDEELKLMAFDIETFEYHEGDEFGKGIIMISYADEEARVITWKNIDLPY	180
Db	121	lidskglpmegeelkmlafdiqctfilyhegeefaeagrimisyadeagrvltwknidpy	180
QY	181	VDVYNEBEMKIRFQOIRREKRODVLITLYNGDNPLYLIRIKRAEKGTYLLIGRKHRE	240
Db	181	vdvystemkrlftrvkekropvlltlyngndifaylkrcekglyntfalgtrgs--e	233
QY	241	PRHMGDSFAVEIKGRHIFDLFPVRRTINLPTTLEAVEAYEALGKTRKSILGAEELAI	300
Db	239	pkigmgtrfavevkgrihfdlpyrlrttlnlpytlaevyavfsgpkvkevyaeeltpra	291
QY	301	WETESMKKLAQYSHEDARATYELCKEFPFMAELAKILGOSWMDVSSSGNGLVEMLL	360
Db	299	wetgenlevarysnedakvlyelgkellpmeaglsrtlqgsllwvssfgnltvemll	356
QY	361	RAYARNEELAPKPDDEEYRRRLRTTYIGGVKPEBERGLMENTITVLDRCLYPSITVHN	420
Db	359	rkeyemelnaphkpekelarr-rqsyagvykeperglwenvilyldtrsltypsllthn	411
QY	421	VSPDTLEBGCANDVAPIVGKFCCKDPGRLPSILGELITMRDIFKKKKKATDPLEKK	480
Db	418	vspdtlmgcekydvaqvyhrrfckdipgrfipslllgelgleorgklkkkmatdpllerk	477


```

FT      Domain      308..318
FT      /note= "Exo 3 region"
PN      EP822256-A2.
XX
XX
PD      04-FEB-1998.
XX
XX      24-JUL-1997; 97EP-0112760.
XX
XX      30-JUL-1996; 96JP-0200446.
XX      29-JUL-1996; 96JP-0198911.
XX
XX      (TOYM ) TOYO BOSEKI KK.
PI      Imanaka T, Kamimura H, Kawakami B, Kawamura Y, Kitabayashi M,
PI      Komatsubara H, Takagi M;
XX
DR      WPI: 1998-102909/10.
XX
XX      Modified thermostable DNA polymerase - with reduced 3'-5'
XX      exonuclease activity
XX
XX      Claim 44; Page -: 54pp; English.
XX
XX      The invention claims a mutated thermostable KOD DNA polymerase of the
XX      present sequence. It has 48% of the 3'-5' exonuclease activity when
XX      compared with the wild-type KOD DNA polymerase. It also has a DNA
XX      extension rate of at least 120 bases/sec, while retaining at least 60%
XX      of its activity after heating at 95 deg. C for 6 hours. The invention
XX      also claims other mutated versions of the polymerase (AAW48449-W48461).
XX      Aspartate 141 and glutamate 143 are known to be essential for the
XX      exonuclease activity. The invention shows that efficiency in DNA
XX      amplification can be achieved by using a mixture of two different DNA
XX      polymerases which are almost identical to each other with respect to
XX      thermostability and DNA extension rate but are different in their 3'-5'
XX      exonuclease activity. The mixture may contain a mutant polymerase
XX      (AAW48449-W48453) having 0-5% of the 3'-5' exonuclease activity of a
XX      native KOD polymerase used with a second polymerase which may be the
XX      wild-type KOD polymerase or the present one.
XX      N.B. The present sequence is not given in the specification, but is made
XX      up using the given wild-type KOD polymerase sequence and the information
XX      provided in the claims section.
XX
XX      Sequence 774 AA.
XX
XX      Query Match      80.0%; Score 3220.5; DB 19; Length 774;
XX      Best Local Similarity 77.4%; Pred. No. 6.3e-243;
XX      Matches 598; Conservative 82; Mismatches 90; Indels 3; Gaps 2;
XX
XX      1 MFDPTYITKDGKPIIRIKKENGKFEKILDPHFQPIYIALKDDSAIDELKAIKGERHG 60
XX      1 MLDITGLTEGDKPVRIRIKKENGKFEKILDPHFQPIYIALKDDSAIDELKAIKGERHG 60
XX      61 KIIVRVDAVKVKKKFGVWKKLFEHPDPAALRGIRHPAVIYDIYEDIPAKRY 120
XX      61 LVVTVRVKVEKQKIFGVWKKLFEHPDPAALRGIRHPAVIYDIYEDIPAKRY 120
XX      121 LIDKGLIPMEGDEELKMFADIEFYHGEDFEGKGEIIMISYADEBEARVITWKNIDLPY 180
XX      121 LIDKGLIPMEGDEELKMFADIEFYHGEDFEGKGEIIMISYADEBEARVITWKNIDLPY 180
XX      121 LIDKGLIPMEGDEELKMFADIEFYHGEDFEGKGEIIMISYADEBEARVITWKNIDLPY 180
XX      181 VDVVSNEREMKIRFOVIVKEDPVLIYNGDNFDLPYIKRAEKIGVTLLEGROKHEPE 240
XX      181 VDVVSNEREMKIRFOVIVKEDPVLIYNGDNFDLPYIKRAEKIGVTLLEGROKHEPE 240
XX      181 VDVVSNEREMKIRFOVIVKEDPVLIYNGDNFDLPYIKRAEKIGVTLLEGROKHEPE 240
XX      241 PRIRHMGDSFAVEIKGIRHFDLPVVRRTINLPYTLVLEAVYEVGKTSKIGAEETIAI 300
XX      241 PRIRHMGDSFAVEIKGIRHFDLPVVRRTINLPYTLVLEAVYEVGKTSKIGAEETIAI 300
XX      239 PKIQTGMDTFEVEVKGRIFLIDLPVIRITINLPYTLVLEAVYEVGKTSKIGAEETIAI 298
XX      301 WETESMKRLAQSMEDARATYELGKEFPMEAEKLAIKGQSVWDVSRSTGNLVEWYLL 360
XX      299 WETGENLEVARYSMEDAKVYELGKEFIPMEAGSLRILGQSLWDSRSTGNLVEWYLL 358

```

```

OY      361 RYAEENNELAPNKKPDEEYRRRLRTTYLGYYKEPERGLMENTITYLDFRCLYPSIIVTHN 420
OY      359 RkayerneLapnKpkdekElarr-rqsyeggykEperglmentIyldfrclypsIilthn 417
DB
OY      421 VSPDITLEREGCKNYDVAPLVGVKFCDFPFGFIPSIITIGELITMRQELKKMKATIDPIEK 480
DB      418 VSPDITLEREGCKNYDVAPLVGVKFCDFPFGFIPSIITIGELITMRQELKKMKATIDPIEK 477
OY      481 MLDYORAVKLANSYGYMGYPKARWYSKEAESYTAGRHVYEMTITEIEKKEFGKVL 540
DB      478 IIdygrakIlanSYgyygaratwykEcaesytavagreyIemtlkeekygfKvi 537
OY      541 YADTDFGYATIPGCEKPEETIKRAKEFLKYINSKLPGLLELEGFYLNGCFYAKKRYAVI 600
DB      538 yadtdgfYatIpGceKpeEtIkRaKeFlKyInSkLPglLeLeGfYlNgCfYakKryAvI 597
OY      601 DEGRITTTGLEVRRDMSIEKIQAKVLEALKEDSVKAEVLYKDVVEIATKYQVPL 660
DB      598 deegritTtgleVrrdmsIEkIQakVleAlkEdSVkAEvLYkDVVEIATkyQvPl 657
OY      661 EKLVIHQITKDLSEYKAIGPVAIAKRLAAKGIKVRPCTIISYIVLRSGKISDRVILL 720
DB      658 ekLvhEqItRdLkYkagVhvaVakRlaargvKlRptvIsyIVlRsgKISdrVlP 717
OY      721 SEYDPKRRHYPDPYIENOVLAIVLILFAFGYRKREDLKYOSSKOYGLDAMUK 773
DB      718 deFpRkRhYpDpYIeNOvLaiVlIlfAfgYrKRedLkyOssKOyGLdAmUk 770

RESULT 14
AAW48450
ID      AAW48450 standard; protein; 774 AA.
XX
XX      AAW48450;
AC
XX
XX      15-JUL-1998 (first entry)
DT
XX
XX      Mutant KOD DNA polymerase (143BA).
DE
XX
XX      KOD DNA polymerase; 3'-5' exonuclease activity; amplification.
KW
XX
XX      Pyrococcus sp.
OS
XX
XX      Synthetic.
XX
XX      Key      Location/Qualifiers
XX      FT      Misc-difference 143
XX      FT      /note= "Changed from Glu in wild-type
XX      FT      to Ala in mutant"
XX      FT      Domain      137..146
XX      FT      /note= "Exo 1 region"
XX      FT      Domain      206..222
XX      FT      /note= "Exo 2 region"
XX      FT      Domain      308..318
XX      FT      /note= "Exo 3 region"
XX
XX      EP822256-A2.
XX
XX      04-FEB-1998.
XX
XX      24-JUL-1997; 97EP-0112760.
XX
XX      30-JUL-1996; 96JP-0200446.
XX      29-JUL-1996; 96JP-0198911.
XX
XX      (TOYM ) TOYO BOSEKI KK.
PI      Imanaka T, Kamimura H, Kawakami B, Kawamura Y, Kitabayashi M;
PI      Komatsubara H, Takagi M;
XX
XX      WPI: 1998-102909/10.
XX
XX      Modified thermostable DNA polymerase - with reduced 3'-5'

```

PT exonuclease activity
 XX Claim 9; Page -: 54pp; English.
 PS
 XX The invention claims a mutated thermostable KOD DNA polymerase of the
 CC present sequence. It has undetectable 3'-5' exonuclease activity when
 CC compared with the wild-type KOD DNA polymerase. It also has a DNA
 CC extension rate of at least 30 bases/sec, while retaining at least 60% of
 CC its activity after heating at 95 deg. C for 6 hours. The invention also
 CC claims other mutated versions of the polymerase (AAW48449-W48461). The
 CC invention shows that efficiency in DNA amplification can be achieved by
 CC using a mixture of two different DNA polymerases which are almost
 CC identical to each other with respect to thermostability and DNA
 CC extension rate but are different in their 3'-5' exonuclease activity.
 CC The mixture may contain the present polymerase with a second one which
 CC may be the wild-type KOD polymerase or a mutant polymerase
 CC (AAW48454-W48461) having 6-100% of the 3'-5' exonuclease activity of a
 CC native KOD polymerase. N.B. The present sequence is not given in the
 CC specification, but is made up using the given wild-type KOD polymerase
 CC sequence and the information provided in the claims section.
 XX
 XX Sequence 774 AA:
 50
 Query Match 80.0%; Score 3219.5; DB 19; Length 774;
 Best Local Similarity 77.4%; Pred. No. 7.5e-243;
 Matches 598; Conservative 82; Mismatches 90; Indels 3; Gaps 2;

DB 658 eklyheqitrdldkyatgphavakrilaargykkiprgvctislyivkgsgrldgripf 717
 QY 721 SEYDPKKKKDDPDYIENOVLPVAVLRILEAFGRKKEDLKYOSKQVGLDAMLK 773
 DB 718 defqptkhydaeylienyvlpaverillrafykedlrygktrvglsawlk 770
 RESULT 15
 AAW48454
 ID AAW48454 standard; protein; 774 AA.
 XX
 XX AAW48454;
 AC
 XX 15-JUL-1998 (first entry)
 DT
 XX Mutant KOD DNA polymerase (142IN).
 DE
 XX KOD DNA polymerase; 3'-5' exonuclease activity; amplification.
 KM
 XX Pyrococcus sp.
 OS
 XX Synthetic.
 XX
 XX Key Location/Qualifiers
 FH Misc-difference 142
 FT /note- "Changed from Ile in wild-type
 FT to Asn in mutant"
 FT Domain 137..146
 FT /note- "Exo 1 region"
 FT Domain 206..222
 FT /note- "Exo 2 region"
 FT Domain 308..318
 FT /note- "Exo 3 region"
 FT
 PN EP822256-A2.
 XX
 XX 04-FEB-1998.
 PD 24-JUL-1997; 97EP-0112760.
 PE
 XX 30-JUL-1996; 96JP-0200446.
 PR 29-JUL-1996; 96JP-0198911.
 XX
 XX (TOYM) TOYO BOSEKI KK.
 PA
 PI Imanaka T, Kamimura H, Kawakami B, Kawamura Y, Kitabayashi M;
 PI Komatsu Arah, Takagi M;
 DR WPI; 1998-102909/10.
 XX
 XX Modified thermostable DNA polymerase - with reduced 3'-5'
 PT exonuclease activity
 PT
 PS Claim 40; Page -: 54pp; English.
 XX
 XX The invention claims a mutated thermostable KOD DNA polymerase of the
 CC present sequence. It has 95% of the 3'-5' exonuclease activity when
 CC compared with the wild-type KOD DNA polymerase. It also has a DNA
 CC extension rate of at least 120 bases/sec, while retaining at least 60%
 CC of its activity after heating at 95 deg. C for 6 hours. The invention
 CC also claims other mutated versions of the polymerase (AAW48449-W48461).
 CC Aspartate 141 and glutamate 143 are known to be essential for the
 CC exonuclease activity. The invention shows that efficiency in DNA
 CC amplification can be achieved by using a mixture of two different DNA
 CC polymerases which are almost identical to each other with respect to
 CC thermostability and DNA extension rate but are different in their 3'-5'
 CC exonuclease activity. The mixture may contain a mutant polymerase
 CC (AAW48449-W48453) having 0-5% of the 3'-5' exonuclease activity of a
 CC native KOD polymerase used with a second polymerase which may be the
 CC wild-type KOD polymerase or the present one.
 CC N.B. The present sequence is not given in the specification, but is made
 CC up using the given wild-type KOD polymerase sequence and the information
 CC provided in the claims section.

XX Sequence 774 AA:

Query Match 79.9%: Score 3218.5; DB 19; Length 774;
 Best Local Similarity 77.4%: Pred. No. 9e-243;
 Matches 598; Conservative 82; Mismatches 90; Indels 3; Gaps 2;

```

QY 1 MIEDDYITKDGKPIIRFKKENGFEKILDPHFQPIYVALKDDSAIDIRAKNGERNG 60
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 1 mldtdytleedgkpvrltkkengefkleydrctfeyfyalikddsaieevkkitaerng 60
QY 61 KIVRVDAVKVKKKFLGRVVEWVKLIFEHPQDVPAALRGKIREHPAVIDIYEDIPAKRY 120
  :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 61 lvtvkrvekvqkflgrvvewklyflhpqdvpaalrckirehgauidiyedipfakry 120
QY 121 LIDKGLIPMEGDEELKLMFADIETFYHGEDEFGKEIIMISYADEPEARVITMKNIDLPY 180
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 121 lldkglvpmegdeelkmlafdneclyhegefaegpilmisyadeegarvltmknvdlpy 180
QY 181 VDVSNREMIKRFVOIVREKDPDVLITYGDNFDLPYLKRAEKLGVTLGLGRKEHPE 240
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 181 vdvstereemlkrflrvvkekdpdvlitygdnfdfaylkrceklglnfalgrdgs--e 238
QY 241 PRIHRGDSFAVEIKRHFEDLPVVRKRTINLPTYTLEAVYEAVALGKTSKLGAEETAAI 300
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 239 pkiqrmgdfavevkgvrlfdlpyvlrtlnlpytleavyaavfqpkekvaaeelpa 298
QY 301 WETESMKLAQYSMEDARATYELCKEFPMEAEI LAKIGOSVMDVSRSTGNLVEMWYL 360
  |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 299 wetgenleratysmedakvtyelqkelfpmaeqslrlqgslwdvstrssignlwevfl 358
QY 361 RYAYENELAPKPKDEEYRRRLRTTYLGQYKKEPERGLMENITYLDFRCLYPSIIVTHN 420
  |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 359 rkayernelapnkpkdekelarr-rqsyeggykpeperglwenlyldfrslpsyllthn 417
QY 421 VSPDTEREGCKNYDAPIVGKFCDFPGFIPISITGELITMROEIKKKKATIDPIEKK 480
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 418 vspdtlnregckeydvapgvgnr-fckdipgflpsllgdlleerqklkkmkatidpietk 477
QY 481 MLDYRQAVKLANSYGYGYGPKARWYSKCAESVTANGRYIEMTKEIEKEGFKVL 540
  :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 478 lldyqraikllansgygyygararwyckecaesvtangreylmtlkeiekygfkvl 537
QY 541 YADTGFVATTIGEEKPETKKKAKKFLKTYNSKLPGLLELEYEGFTLRGFFVAKKRYAVI 600
  |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 538 yadtgffatclpdaaetvkkameflnylnaklpaleleyegfykrgffvltkkkyavl 597
QY 601 DEGRITTRGLEVVRDMSKIAKETOAKVLEALIKEDSVKAEIYKDYVEEIAKYOVL 660
  |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 598 deegklcttrgleivtrtdwseleketqarvleallkdgdvekavrlvkevtelkskyevpp 657
QY 661 EKLVIHQITKDLSEYKAIGPVATAKRLAAGKIVRPCTIISYIVLRSGKISDRVIL 720
  |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 658 eklvlnheqitrdlkdykatgphavakrlaargvklrpgtvsyvlksgslgdraipf 717
QY 721 SEYDPKKKKYVDPDYIENOVLPVLRILEAFGRKKEDLKQSSKOYGLDAMLK 773
  |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 718 defdpkpkhkydaeylengvlpaverillrafgykrkedlrygktrgygsawl 770

```

Search completed: May 28, 2002, 14:18:46
 Job time: 104 sec

Q9H98 PRELIMINARY: PRT: 759 AA.
 AC Q9H98
 DT 01-MAR-2001 (Tremblrel. 16, Created)
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE DNA POLYMERASE (FRAGMENT).
 GN POL.
 OS Pyrococcus sp. (strain ST700).
 OC Archaea: Euryarchaeota: Thermococcales; Thermococcaceae; Pyrococcus.
 NCBI_Taxid=69015;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ST700;
 RA Querellou J.J.E., Cambon M.A., Lesongeur F., Barbier G.;
 RT "Thermococcales taxonomy and phylogeny based on the comparative use of
 RT 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
 RT genes";
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -> N
 CC PYROPHOSPHATE + DNA(N).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 DR EMBL: AJ250332; CAC12847.1; -.
 DR HSSP: P56689; ITGO.
 DR InterPro: IPR002064; DNA_POL_B.
 DR InterPro: IPR001648; Ribosomal_S18.
 DR Pfam: PF00136; DNA_POL_B; 1.
 DR Pfam: PF03104; DNA_POL_B_exo; 1.
 DR PRINTS: PR00106; DNA_POLB.
 DR SMART: SM00486; POLBc; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B; UNKNOWN_1.
 DR DNA replication; DNA-binding; DNA-directed DNA polymerase.
 FT NON_TER 759 759
 SQ SEQUENCE 759 AA: 88431 MW: A8459B6AD8F3AF3 CAC64;

Query Match 76.5%; Score 3081; DB 1; Length 759;
 Best Local Similarity 74.7%; Pred. No. 5.3e-173;
 Matches 569; Conservative 99; Mismatches 90; Indels 4; Gaps 3;

QY 1 MFFDDYITKDGKPIIRIFKKEGFEKIELDHPFQPIYIALLKDSADIDEIKATIGERG 60
 DB 1 MLDDADYITENKPIIRIFKKEGFEKIEYDNPFPYIALLKDSADIDVKKITSERG 60
 QY 61 KIVRVADVAVKKKFLGRDVEYWKLFEPHODVPALRGKIREHPYVIDIYEDIFPAKY 120
 DB 61 KIVRVADIEKSKKFLGRDIEYWKLFEPHODVPALRGKIREHPYVIDIYEDIFPAKY 120
 QY 121 LIDKGLIPMEGDEELKMAFDIETFYHEGDEFGKEIIMISYADEEGARVITWKNIDLPY 180
 DB 121 LIDKGLIPMEGDEELKMAFDIETFYHEGDEFGKEIIMISYADEEGARVITWKNIDLPY 180
 QY 181 VDVVSNERNEMIKRFQVYERKDPVLITNGDNFDLPYIKRAEKLGYILLGGRKEHE 240
 DB 181 VDVVSNERNEMIKRLVIRREKDPDIIITYNGDNFDLPYIKRAEKLGYILLGGRKEHE 240
 QY 241 PIKHMGDSFAVEIKGRIHFDLPVVRTINLPYTLAEVAVYAGKTSKLGAEIAT 300
 DB 241 PIKHMGDSFAVEIKGRIHFDLPVVRTINLPYTLAEVAVYAGKTSKLGAEIAT 300
 QY 239 PMQMLGSLANVEIKGRIHFDLPVIRKRIINLPYTLKRVYAIATIGKREKYPHEIATA 298
 DB 239 PMQMLGSLANVEIKGRIHFDLPVIRKRIINLPYTLKRVYAIATIGKREKYPHEIATA 298
 QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEALAKLIGOSVWVSRSSTGNLVEWYLL 360
 DB 301 WETESMKKLAQYSMEDARATYELGKEFFPMEALAKLIGOSVWVSRSSTGNLVEWYLL 360
 QY 361 RYAVRNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMETIYLDPRCLYPSIITVFN 420
 DB 361 RYAVRNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMETIYLDPRCLYPSIITVFN 420
 QY 421 VSPDTEREGCNVAVPIVGYKFCDFGFIPIISIGELITTKROETKKMKATPIIEK 480
 DB 421 VSPDTEREGCNVAVPIVGYKFCDFGFIPIISIGELITTKROETKKMKATPIIEK 480
 QY 419 VSPDTEREGCNVAVPIVGYKFCDFGFIPIISIGELITTKROETKKMKATPIIEK 478
 DB 419 VSPDTEREGCNVAVPIVGYKFCDFGFIPIISIGELITTKROETKKMKATPIIEK 478
 QY 481 MLDYRORAVKLANSYGYGMYPKARWYSKECAESVTAWGRHYIDMTIKELIEKGFYVL 540
 DB 481 MLDYRORAVKLANSYGYGMYPKARWYSKECAESVTAWGRHYIDMTIKELIEKGFYVL 540

DB 479 ILDYRQRAIKILANSFYGYGAKARWYCKECAESVTAWGRHYIELVRELEER-GEKVL 537
 QY 541 YADTGGFVATITGEGP-ETIKKKAKEFLKYINSKLPGLLEIEGFIYLGFFVAKRVAV 599
 DB 538 YIDTGLVATITGEGP-ETIKKKAKEFLKYINSKLPGLLEIEGFIYLGFFVATITG 597
 QY 600 IDEEGRIITRGLEVVRDSEIAKETQAKVLEAILKEDSVFAVEIVKRVIEIKAYOP 659
 DB 598 IDEEGRIITRGLEVVRDSEIAKETQAKVLEAILKEDSVFAVEIVKRVIEIKAYOP 657
 QY 660 LEKVIYHEQITKDLSEYKAGIPHAIAKRLAKGIVRPGTIISYVLNGSKISDRVIL 719
 DB 658 VEKVIYHEQITKDLSEYKAGIPHAIAKRLAKGIVRPGTIISYVLNGSKISDRVIL 717
 QY 720 LSEYTPKHKHYPDYIENQVLPAYLRILEAGYKREDKYO 761
 DB 718 IEEFDGKKRKYDAEYIENQVLPAYLRILEAGYKREDKYO 759

RESULT 4
 Q9P9N1 PRELIMINARY: PRT: 781 AA.
 AC Q9P9N1
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE DNA POLYMERASE B3.
 OS Sulfurisphaera ohwakuensis.
 OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfurisphaera.
 NCBI_Taxid=69656;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TA-1;
 RX MEDLINE=20450680; PubMed=10997874.
 RA Iwai T., Kurosawa N., Itoh Y.H., Kimura N., Horinouchi T.;
 RT "Sequence analysis of three family B DNA polymerases from the
 RT thermophilic crenarchaeon Sulfurisphaera ohwakuensis";
 RL DNA Res. 7:243-251(2000).
 CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -> N
 CC PYROPHOSPHATE + DNA(N).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 DR EMBL: AB032376; BA033703.1; -.
 DR HSSP: P56689; ITGO.
 DR InterPro: IPR002064; DNA_POL_B.
 DR Pfam: PF00136; DNA_POL_B; 1.
 DR Pfam: PF03104; DNA_POL_B_exo; 2.
 DR PRINTS: PR00106; DNA_POLB.
 DR SMART: SM00486; POLBc; 1.
 DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
 DR DNA replication; DNA-binding; DNA-directed DNA polymerase.
 SQ SEQUENCE 781 AA: 90384 MW: 9B2570EAE30C372C CAC64;

Query Match 29.9%; Score 1204.5; DB 1; Length 781;
 Best Local Similarity 36.6%; Pred. No. 1.1e-62;
 Matches 294; Conservative 141; Mismatches 282; Indels 87; Gaps 21;

QY 2 IDTDYITKDGKPIIRIFKKEGFEKIELDHPFQPIYIALLKDSADIDEIK--AIKGR 58
 DB 7 ILDFSDVVENKPIYIYVWIDKEGNRVVLEKRFYFALVDSDSYNDIEIKREILKLSK 66
 QY 59 HKIYRVADVAVKKKFLGRDVEYWKLFEPHODVPALRGKIREHPYVIDIYEDIFPAK 118
 DB 67 PPSPIITSDV--EEKKYGSPVKVLEIETVITAYVRVYDVEAKIKGASVLEADIRFTW 124
 QY 119 RYLDIKGLIPM-----EGDE-----ELKMAFDIETFYH 147
 DB 125 RKSIDINLKPFVWIEAEVEIKENNRVKKYVELKINKLYEDKIPELKVLAFDIEVY-- 182
 QY 148 EGDDEG-----KGEIIMISYADEEGARVITWKNIDLPYDVVSNERNEMIKRFVQVIREK 202
 DB 183 --NKYGSFNPBRDPVYIIGVMTKEGKQFLADKYD-----DLRAIRFEPINFGVYD 231


```

Db      288 PEKLTLDVAE-YFGVMREERVLPGHKITEYWDPKRPLKYYLDDVSTGLA 346
QY      326 KEFPEAEALAKIGOSVMDVSRSTGNLWEMYLRLVAYERNELAPKPEEERRLRT 385
Db      347 DKLLPILQSSVSGPLDQVAAASVGNRYEMMLKRAYLGEVAPNR-EEREY----- 400
QY      386 TYLGGYKKEPERGLMENTYLDRCFLYPSLIIVTHNVSPDTL---ERECKNDVAIVG 441
Db      401 PKGATVLEPKPGMEYEDVLDFSSMYPINMKYNLSPTDYLPEPEDPPEGVNAPEVG 460
QY      442 YKCKXDPFGPIPSILGELTMRQELKKMKA-TIDPLEKMDLYRQAVLHANSYGYM 500
Db      461 HRRFRSPGFPQVQLSLVSLRAVREAAKYPDSEFILLDERPAKLVMANATGYL 520
QY      501 GYKAMWYSEKCAESYVAMGRHYEMTIKEIEKFGFKVYADTDGFYATIPGKPEPIK 560
Db      521 GWGAMRYKREVAESYAFARALKDVIQ-ARRLGIVVYGGTDSLFFVKKHD----- 573
QY      561 KKAKEFLKYINSLPGLLEVEYEGFYLRGFV-AKKRAYVIDEGRITTRGLEVVRDWS 619
b      574 --VDKLTLYVEEKY--IDIKVDKDYAKVLFTEAKKRYAGLLDRGIDIVGFVVGDS 629
QY      620 EIMKEQAKYLEALKEDSDEKA----VEIVKDYVEEIAKYQVPLEKLYHEQITDLS 675
Db      630 ELAKDVQLRYELILSRDIVEARHGVIKIRIELRLKMYKFNIDLLIWKTLDELDE 689
QY      676 YKAGHPVAIAKRLAKGKIRGPTIISYVLRSKISDRV---ILLSEYDKKHKHYD 732
Db      690 YKAYPPHVAHQILKRGYRVGKGTIGYIVYVAGKEVSEALPYILLDDI-----KKIDI 745
QY      733 DYIENQVLEPAVLRLIEAGCYRKEDK 759
Db      746 DYIERQIIPALRLIAEIVGKESDLK 772

RESULT 8
ID      09HH85      PRELIMINARY;      PRT;      195 AA.
AC      09HH85;
DT      01-MAR-2001 (TREMBLrel. 16, Created)
DT      01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      DNA POLYMERASE (FRAGMENT).
GN      POL.
OS      Thermococcus chiltonophagus.
OC      Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX      NCBI_TaxID=54262;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Querellou J.J.E., Cambon M.A., Lesongeur F., Barbier G.;
RT      "Thermococcales taxonomy and phylogeny based on the comparative use of
RT      16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
RT      genes.";
RL      Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR      EMBL: A250334; CAC12848.1; -.
DR      HSSP: P56689; ITGO.
DR      InterPro: IPR02064; DNA_pol_B.
DR      Pfam: PF03104; DNA_pol_B_exo; 1.
FT      NON_TER      1
FT      NON_TER      195
SQ      SEQUENCE      195 AA; 22453 MW; 23FA457E4EFA8418 CRC64;

```

Query Match 19.2%; Score 774; DB 1; Length 195;
 Best Local Similarity 73.6%; Pred. No. 3.2e-38;
 Matches 15; Conservative 27; Mismatches 23; Indels 2; Gaps 1;

```

QY      122 IDKGLPMGEDELKLMARDIETFYHGDFFGKGLIMISYADEBARAVITMKNIDLPYV 181
Db      1 IDKGLPMGEDELKLLAFDIETLYHGEFFAKGITIMISYADEEGKITVMKKVDPYV 60
QY      182 DVVSNREMIKRVQIVREKDPVLTITNGDNFDLPYLKRAEKLGVTLTLGDKKEHP 241

```

```

Db      61 EVVSSEREMIKRFLKYIREKDPVITTYNGDSFDLPYLKRAERLGIKPLGRDGS--EP 118
QY      242 KIRMGDSFAVELIKGRHFDLPVVRTINLPYTLAEVAVENALGKTSKLGAEIATV 301
Db      119 KMORLGMVAEIKGRHFDLVHVRTINLPYTLAEVAVENALGKPKREKVAVHETAEAV 178
QY      302 ETEESMKKLAQYSMEDA 318
Db      179 ETGKGLERVARYSMEDA 195

```

```

RESULT 9
ID      09NAH1      PRELIMINARY;      PRT;      1443 AA.
AC      09NAH1;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE      Y47D3A.29 PROTEIN.
DE      Y47D3A.29.
GN      Caenorhabditis elegans.
OC      Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC      Rhabditidae; Pelodierinae; Caenorhabditis.
OX      NCBI_TaxID=6239;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Matthews L.;
RL      Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RX      MEDLINE-99069613; PubMed-9851916;
RA      none;
RT      "Genome sequence of the nematode C.elegans: A platform for
RT      investigating biology.";
RT      Science 282:2012-2018(1998)
CC      -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -> N
CC      PYROPHOSPHATE + DNA(N).
CC      -1- MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC      ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC      DIFFERENT REACTIONS OF DNA SYNTHESIS (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR      EMBL: AL117202; CAB97239.1; -.
DR      InterPro: IPR002064; DNA_pol_B.
DR      PRINTS: PR00106; DNAPOLB.
DR      SMART: SM00486; POLBc.1.
DR      PROSITE: PS00116; DNA_POLYMERASE_B.1.
DR      DNA replication; DNA-binding; DNA-directed DNA polymerase.
SQ      SEQUENCE      1443 AA; 163380 MW; EF4B2E9E0A50BB8 CRC64;

```

Query Match 15.7%; Score 631.5; DB 5; Length 1443;
 Best Local Similarity 26.0%; Pred. No. 1.3e-28;
 Matches 206; Conservative 155; Mismatches 301; Indels 131; Gaps 27;

```

QY      63 VRYVDAVKKKPKFLGRDVEYWKLIFFHPDVPALRGKIREHPAVIDIYEYDIPAF---- 118
Db      462 VTLERLLIEKKPMG-----PQIMILVNSDPKATSNQ 495
QY      119 -----RYLIDKGLPMGEDELKLMARDIETFYHGDFFGKGLIMISY-- 162
Db      496 KYEEVDEMERNIKYLESSDAPP---PTIKLLALNVTTLLDRKD---NEICMSMLF 548
QY      163 -----ADEEAR-----VITMKNIDLPY-----DVVSNREMIKRF 194
Db      549 NPKCDLAHPSGDSKDLKPKCVIYKPPGSGSPYIQRLKETLSKRYKSVANAKALLTLF 608
QY      195 VQIVREKDPVLTITNGDNFDLPYLKRAEKLGVTLTLGDKKEHPKTIHRMGDSFAVE- 253
Db      609 LAMINDEDPMIVGHD-LSATIALIVSRLEKTLKPMWSRISRLKRSINIGIKIGHSQGW 667
QY      254 --IKGRINHEDFPVVARTINLPYTLAEVAVENALGKTSKLGAEIATVETESMKKLA 311

```

```

Db 668 ELTAGRMILDSKLAMELVSKSPDLTETSOILGAERREIYANEIPOLYVDSKDLISLI 727
Qy 312 QYMEDARATYELGKEF--FPMEALAKLIGSVWDVSRSSNGNLVE--WTLLRATYARN 367
Db 728 NWSNADLLSLRIVYVRLAALPLIYIOISIVGFI--ASRTIMGGAERNEYLLHAFEKA 784
Qy 368 EL-APNKPDEEYRRRL-----RTYLGVEYKEPERGIMEN-ITVLDRCCLYP 413
Db 785 DLIAADKYNSSSEKKKKKEESGDEKSKGAKQYSGGLVEPKGKGLYELLILLDLSNLYP 844
Qy 414 SIYTHANVSPOTLEREGCKNDVAPVIGYKCKPFGFPIPSILGELITMROEIKKKMAT 473
Db 845 SIIOGYNCITLLEYS--KDSDEQLSVPOS--TDIEGVLPREIRKLVCEGRDVKSLMKSE 900
Qy 474 IDPIKKMLDVRORAVKLHANSYGYMGYPKARWYSKCAESVTMGSHYEMTKEEE 533
Db 901 RREAKKKOMDROMALKLITANSMTGCLGFQYSRFAKPLALVITAKGRE-ILMSKDLVE 959
Qy 534 KFGKVLVADTDFGFAVITPGEKPEIKRKAKEFLKYINSKLPGLLEVEYEGFVAFVA 593
Db 960 KMGYSVYVGDTSIMINTNSVDLVAKKLGNKIKKAVN-KCHRLLELDLDGVEKMLLTK 1018
Qy 594 KRRYA--VIDEGRIT---KGLVYVRDWEIANKETQAKVLEAIL-----KEDSVKAV 643
Db 1019 KKKYALVIMPTKVEAKELGLDVRDWSQLAKETGAVVADKILDSLRDEWVSSID 1078
Qy 644 EIVKDVVEIAKYOVPLKELVIEHOITKOLSEYKAIQP--HVAIAKRLAANG-IVRPGT 700
Db 1079 DLKREIRAKLDSGYVPLEMFOISKQLTNRPEQYADYKQCHAAVAQRLANKSKFYLRLND 1138
Qy 701 IISYIVLNG--SGKISDVILLSEY-DPKKHKYDPDYVYIENOVLPVAILTEAFGRKED 757
Db 1139 IVEYVICEDGTNDNPATQARHTEMAENSALKIDLLYLAQIHPVSRIVEPI--ET 1195
Qy 758 LKYOSKOVGIDA 770
Db 1196 DAVRIAENLGLDS 1208

```

```

RESULT 10
Q9HRV9
AC Q9HRV9 PRELIMINARY; PRT; 901 AA.
DT 01-MAR-2001 (TRENBLREL. 16, Created)
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE DNA POLYMERASE B1.
NM POLB1 OR VNG0521G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA NG W.V., Kennedy S.P., Mahaites G.G., Bergquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sirogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenberger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freltas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT Genome sequence of Halobacterium species NRC-1";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE = N
CC PYROPHOSPHATE + DNA(N).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
DR EMBL: AE005004; AAG19049.1;
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam: PF00136; DNA_pol_B.2.
DR Pfam: PF03104; DNA_pol_B-exo; 1.
DR PRINTS; PR00106; DNAPOLB.

```

```

DR SMART; SM00486; POLB; 1.
DR PROSITE; PS00116; DNA POLYMERASE B; 1.
KW Complete proteome, DNA replication; DNA-binding;
KW DNA-directed DNA polymerase.
SQ SEQUENCE 901 AA; 101987 MW; 2134704B7AB5E49D CRC64;

```

```

Query Match 15.5%; Score 626; DB 17; Length 901.
Best Local Similarity 26.7%; Pred. No. 1.4e-28;
Matches 242; Conservative 138; Mismatches 319; Indels 208; Gaps 42;

```

```

Qy 1 MIFPDYITKDG-----KPIIRIFKENGFEKIEIDPH-FQPIY---ALKDSDAIDEI 51
Db 60 MVTQVDY-TIGSGDGRREPLVLFVETRTDDEAHNRVAGFRYFAFATILSEDLTDYD 118
Qy 52 -----KAIRGERHGIYRVYDAVKKKFLGDVEWKLIFHPDVPALCKIREH 103
Db 119 ITGEDGESIRGEALTKI-----FGRT-----PRDV-----GNIRD- 150
Qy 104 PAVIDYEDYDIPKARYLIDK-----GLIPMGDE-----ELKMAF 140
Db 151 --AFDHYEADILFPNRLIDDKDITDGVVPPARRADGALVYHHDEIACEVAADLRVTF 208
Qy 141 DIEFFYHGC-DEFGKEITIMISYADEEARVITW-----KNIDLPY----- 180
Db 209 DIEVDRNGFPEDGEPVCLTSHDSYRDEYVAMLYAPDATVPATVAPGYDPLTDAD 268
Qy 181 --VDVVSNEREMIKRFVQIVREKDPDVLITNGDNFDLPYLIRAKKAGVITLLGRCKEH 238
Db 269 IDVRVFDTEAMHDAFLSYIEDTNPDLVLTGNFDFDAPVILDRIDELDPR---TDHDL 324
Qy 239 PEKPIHRMGDSF-----AVEIKGRHFDLPFVVRRT--INPTYLEAVYEAVALGKTSK 291
Db 325 DSDRLSRVSEVWTSGMGSPNKGKRVFLLAYQTKSELDSYLDVAVGEDELGVCKER 384
Qy 292 LGAEIIAIIWETESMKRLAOLYSMEDARATYELGKE--FPMEALAKLIGSVWDVSR 349
Db 385 YPG-DIGDLM--EDDPERLLEYNLRLDELVEICVEIDRKQISIVAFWIDRARKLVCKLEDA--T 439
Qy 350 STGNLYEWYLLRVAVERELAPNK--PDEEYRRRLRTTYIGGYKEPERGIMENITVLD 407
Db 440 TPGDVADVYVLRKAF-GNVLPSKQGDAAEFEE-----GGAFFDITVREVSVD 490
Qy 408 FRCLYPSIIVHNANSPD-LEREGCKNDVAPVIGYKCKDPFGFPIPSILGELITMROEI 466
Db 491 LKSLTPMSMTVITNASPEKRVDPEDGTYRTPTGVRHRRKPEDGILIRKMBELLTEREE- 549
Qy 467 KKKMATIDPIEKKMLDY-----RQRAVKLHANSYGYMGYPKARWYSKCAESVTAMGR 521
Db 550 --KARRDDEHPDSEDEYERYDROQAAVKVIMNSLYGVGMDRFLYDRAMSAGVSTNR 606
Qy 522 HYIENTIKEIEKRGFKLVADTDFGFAVITPGE--KPETIKK-----AKE 565
Db 607 EVIDFT--EQAAEEFGYEVAVGDTDSVMLELGDMTKAAIAIESPIDEDHINAAYDEFARE 665
Qy 566 FLKYINSKLPGLLEVEYGFILGPFVA--KKRYA--VIDEGR-----ITRGLEVVARD 617
Db 666 QLHADDNR-----FOIEFKLI--RRFQAGKKRRAGHIVMKDEKVDVDVDTITGFEYQSD 720
Qy 618 WSEIAKETQAKVLEALIKEDSEVAKAVELVQDVVEIARVQVPLKELVITHEOITDLSYK 677
Db 721 IAPITKRVQKEVIDLVVNEGVDVAVEEYVHGVIEEFGCGDADLDIGIPGIGKRLDYE 780
Qy 678 AIGHVAIAKRLAKKAGIKVRCGTIISYVLGS-----GKI----- 713
Db 781 -----TDTAQVRGAKYANLLGTNFD-----RGSKRRLYLAKVHDFWGHVEGQIDTS 831
Qy 714 -SDRV-----TILSEYD--PKKHYPDPDYIENOVLPVAILTEAFGRKEDLKYQ 761
Db 832 TSDRYRFRRPDPVICYEYLAQOIIDTFVYVDKMLDKLKPIERILFALDLSMDEVK-S 890
Qy 762 SSKOVGL 768

```

Db 891 GGTGTCL 897

RESULT 11

ID 09JHRO PRELIMINARY; PRT: 796 AA.

AC 09JHRO; 01-MAR-2001 (Tremblrel, 16, Created)

DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)

DE DNA POLYMERASE (POLB), LARGE CHAIN RELATED PROTEIN.

GN TA0907.

OS Thermoplasma acidophilum.

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;

OC Thermoplasma.

OX NCBI_TaxID=2303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;

RX MEDLINE=20479972; PubMed=11029001;

RA Kuep A., Graml W., Santos-Martinez M.-L., Korteke K.K., Volker C.,
Mewe H.-W., Fishman D., Stocker S., Lupas A.N., Baumeister W.;RT "The genome sequence of the thermophilic scavenger Thermoplasma
acidophilum.";

RT Nature 407:508-513(2000).

CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE - N
PYROPHOSPHATE + DNA(N).

CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

CC EMBL: A445065; CAC12036.1; -

CC Interpro: IPR002064; DNA_pol_B.

CC Interpro: IPR000719; Euk_pkinase.

CC pfam: PF003104; DNA_pol_B_1.

CC pfam: PF003104; DNA_pol_B_exo; 1.

CC PRINTS: PR00106; DNAPOLB.

CC SMART: SM00486; POLBC.1.

CC PROSITE: PS00116; DNA_POLYMERASE_B.1.

CC PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_1.

CC PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

CC Complete proteome; DNA replication; DNA-binding;

CC DNA-directed DNA polymerase.

SQ SEQUENCE 796 AA; 91210 MW; DBCDB0F1463EB929 CRC64;

Query Match

Best Local Similarity 25.5%; Score 623; DB 17; Length 796;

Matches 211; Conservative 161; Mismatches 314; Indels 134; Gaps 31;

16 IRF-KKNGEFKIEDPHOPITVALKDDSAIDELKAKGRHGKIVAVAKKK 74

19 VELLEGRTEESVAALYFGFRP-YDVPEDEAY-LKVIQND-----PEFYKMDK 67

75 FL---GRDVEVMKLFEPHODVPALRGKIREHVAIDIVYEDIPAKRYLIDGL----- 126

68 RLMIKRGYEVKRIYIRSPKAVPEYREK-----CPREVLADIPFHRIYIDLDSGVK 122

127 -----IPMEGDE-----ELKLMADIEFYHEGDEFGKGETIMS 161

123 IIGENISDRSTFTDIIVIRADRIENVDENPMKVLSEFVENEINRENEVDYKILVIG 182

162 YADEEERAVITKKNIDLPYVDVSNREMKKRVQVIREDDPVLLTYNGNDLPLIK 221

183 YVSFSGKVTG-----SLSGEODILIRSVLDIRADDPVITGYNDIGDIPYIK 234

222 RAKLGLVTLILGDRKHEPKIRMGDSFAVEIKGRHLPFVVRITINLPYTLVAVY 281

235 RMDRYGKLEIGDGSIP-----RIMKQF-WRVHGRILSDTMSVKRILHPKHESLDIYA 289

282 EAVLGKTKSLGAELIAIWEITESMKLAOYSMEDARATYELGKEFFPME-----AELA 336

290 NMLGSGKONIDRLHLEDEW--KKRREVIAYCIKADADLIRFEKLMVNNRLMYSVT 347

337 KLIGQVMDVSNKSTGNLYEWYLLRVAVERNELAPNKPDEEYRRRLRTTYL-GGYVKEP 395

Db 348 KL---PLDDVANGTSNYVDSTILIRAADRENIGVPMNOHE-----IKTEELIOGGYVHSI 398

QY 396 ERGLMENITVLDPRCLPSITVTHNVSPDTLEREGCKNVADIVGKFC--KDPFGFT 453

Db 399 GAGLYSNVYLDKSMSPSITIKYNVCFITLDPNG-----EILSPNGVRLSPKKGLIP 454

QY 454 SILGELITMRCEIKKKKATIDPIEKMLDYPORAVKLANSYGYGPKARWSKECA 513

Db 455 RILOELMADREDEVRKMKAKAKTDEREFDGLNALKVLMNTFYGLASSFYFTHKIG 514

QY 514 ESTVANGRIHYEMTKEIEKFGKVLVADTDCGFVATIGEEKETIKKAKEFLKTYNSK 573

Db 515 SAITAFARERIKIISTLESAN-YRVYIGPTDSVFEESGASNEDAIKRRKDLSERL-SR 572

QY 574 LPGL---LELEYGFYLRGFEVAKKRA-----VIDEGRITRGLVVRDMSEIAKET 625

Db 573 EGGITLDFQWVLDPFPSHG---AKKRIAGCVPPDMKGEIVYKGYEVRTDSFDLSQA 629

QY 626 QAKVLEAILKEDSEKAVEIKVDVEEI--AKYQVPLEKLVHEQITKDLSEYKAGPHV 683

Db 630 LSKVIDFILNRD-IDGAIKADLDIKKVRAGDPSIDISLVISRTV-KDQSYRANQESL 687

QY 684 A---IAKRLAAKIKYRPGITISYIVLRGSGKISDKVILLSEDPKKH-----KVDPOY 734

Db 688 ANIRAAKRLIDMGETFVPGKVSIVTNGKR-----TPQEEPIYIGRDLNVKPDMDY 740

QY 735 YIE--NOVLPAVLR-----LEAFGRKEDLK 759

Db 741 YARLSETLGRVIVERTLDGGOKTASLESFSSSSDEAK 780

RESULT 12

ID 097AH3 PRELIMINARY; PRT: 800 AA.

AC 097AH3; 01-OCT-2001 (Tremblrel, 18, Created)

DT 01-OCT-2001 (Tremblrel, 18, Last sequence update)

DE DNA POLYMERASE.

GN TVG0859451.

OS Thermoplasma volcanium.

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;

OC Thermoplasma.

OX NCBI_TaxID=50339;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=GS51 / DSM 4299 / JCM 9571;

RX MEDLINE=20570466; PubMed=11121031;

RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
Kawashima T., Watanabe K., Yamazaki M., Kanehori K., Kawamoto T.,

RA Nunoshima T., Yamamoto Y., Aramaki H., Makino K., Suzuki M.;

RT "Archaeal adaptation to higher temperatures revealed by genomic
sequence of Thermoplasma volcanium.";

RT Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).

RL EMBL: AP000994; BAB59979.1; -

DR Interpro: IPR002064; DNA_pol_B.

DR Interpro: IPR000719; Euk_pkinase.

DR pfam: PF00136; DNA_pol_B_1.

DR pfam: PF00136; DNA_pol_B_exo; 1.

DR PRINTS: PR00106; DNAPOLB.

DR SMART: SM00486; POLBC.1.

DR PROSITE: PS00116; DNA_POLYMERASE_B.1.

DR PROSITE: PS00933; FGGY_KINASES_1; UNKNOWN_1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.

DR Complete proteome.

SQ SEQUENCE 800 AA; 91833 MW; EA9640C12B134929 CRC64;

Query Match

Best Local Similarity 25.9%; Score 622.5; DB 17; Length 800;

Matches 215; Conservative 165; Mismatches 306; Indels 145; Gaps 33;

```

OY 9 TKGKPIIRIFKENGEEFIELDPHF-----QPIYALLKDDSAI---DEIKAIKGEH 59
DB 27 SREGSVSALE-----FGKPYFDVPEPNEILSSIRNDEVEYKEDKVLWNGSMH 78
OY 60 KRIYVAVDAVKKKFLGHDEVEWMLIFEPDVPALRGKIREHPAVIDIYEYDIPAKR 119
DB 79 -----NKKRIYKSPKVPPEYR---RRCP--FEVLAADIPFHRH 112
OY 120 YLIDKGL---IMEGDE-----ELKMAPIETFFYHKG 149
DB 113 FYLDLDGACVRIGEDISGTEGNETTDLVLKIDKIEINIPDNVNLKVLSPFVENEINRE 172
OY 150 DEFGGEIIMISYADEEBEARITMKINIDLPYDVVSNEREMIKRFOVJIREDDPVLTATY 209
DB 113 NVEDGKILVIGYS-----IMFGDIQ--KKEIGHDEKEILIRRIDILRADDPVITGY 224
OY 210 NGDNFDLPYLKRAKLGVTLLGRDKHEPEPKIRHMGDSFAVEIKGRIFHDLFPVRRY 269
DB 225 NIDGYDMLVIGRMEQYGIHNLNIGRGSVP---RIMDQF--WRVHGRILSDTMNNVKKI 279
OY 270 INLPYITLAEVAVLGKTKSKLGAELTAIETESMKKLAQYSMEDARATYELGKEPF 329
DB 280 IHPKHESLDYIAMKLLGEGKDSIDRLNIEAEWQKROD---EVIISYCIKADLTLLRIFEKLR 337
OY 330 PME-----AELAKLIGOSVMDVSRSTGVLVEMYLLEAVYERNEIAPN-----KPEDEBY 379
DB 338 VLERLMFKSTYKL---PLDDVANAAGTSYVDSILIRAADRENIGVPMANHLKDELD 393
OY 380 RRRLRTYIGYVYKPERKGLMENTYLDPRCLPSIITVHNVSPTLREGECKNYDAPI 439
DB 394 -----GGYVHTIGAGLYSNVIVLDPFKSMYPSMIIRKYNCFITLDPKG---EIVSP 440
OY 440 VGYKFC--KDFGFIPTSIGELITMRQELKKMKATIDPIEKMMIDYQRAVKLANSTY 497
DB 441 TGIKRLSPKKKGLPIRLIQELMADRDVKKRKKAKSEDELFYDGIQNALIKVLMNFY 500
OY 498 GYMGYPAKRWYSKECAESVTANGRYIEMTIKEIEKEFGKVLVADTDFGYATIGEKPE 557
DB 501 GYLAASFYRTPDKIGSATATAFARETIKHI--DYLESSHRITYGDTDSVFESEGVKRE 559
OY 558 TIKKAKAEFLKYNSKLPLG---LELEYEGFLRGFEVAKKRYA---VIDEGRITTR 609
DB 560 DAIKTKELISQKL--SEEBGLTLDFOVNLDPFESHG---AKKRYAGCYVPEDMKGEIITK 615
OY 610 GLEVVRRQMSLEIAKETQAKVLEAILKEDSVKAVEIVDVEI--AKYQVLEKLVTHE 667
DB 616 GYEVARTDSFDLQSTALSKVIDFILLDRD--VOGAINVADLVKRVNKGPSIDIESLVISR 674
OY 668 QTTKDLSEKKAIGPVA---IAKRLAKGIRKVPCTIISYIVLRSGKISDRVILLSEYD 724
DB 675 TV-KPFSSKANTDSLANTRAKRLIERGETYFVPGMKVSWIYT--NGKKTPEVEPEYIYG 731
OY 735 PK-KKRYDPDYIENOVLPVAVLILAEFGYRREDLKYSSKQVGDAMLWK 774
DB 732 SKLESKPDMDYAK--RLSETLNKVIDVF--RKD--MVGQNRISLDSFSK 777

```

RESULT 13

```

OY 0971C7 PRELIMINARY: PRT: 876 AA.
AC 0971C7;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE PUTATIVE DNA-DIRECTED DNA POLYMERASE.
GN ST1426.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota: Sulfolobales: Sulfolobaceae; Sulfolobus.
OY NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;

```

```

RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankaï A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamaya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagis M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.,
RT "Complete genome sequence of an aerobic thermocacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.",
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000986; BAB6493.1;
KW DNA-directed DNA polymerase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 876 AA; 101584 MW; C013A0FB0831F96 CRC64;

Query Match 15.3%; Score 616.5; DB 17; Length 876;
Best Local Similarity 26.6%; Pred. No. 4.8e-28;
Matches 231; Conservative 140; Mismatches 316; Indels 181; Gaps 35;

OY 2 IFDTYITKDGPIIRIFKENGEEFIELD--PHQPIYALLKDDSAIDEIKAIKGERHG 60
DB 53 LLOYVDGKSKAVCKLDTKETKIYILDNTGKPYF-----LTDIDPEKVN 100
OY 61 KIVRVY-----DAVKKKKFLGRDVEWMLIFEPDVPALRGKIREHPAVIDIYE 111
DB 101 KIPKVRDPSFDHLETVIKI--DPYSGNKIKLTKIYVKDPLAVRRMRNSVPR-----AYE 153
OY 112 YDIPPAKRLYDKLIP-----MEGDE----- 133
DB 154 AHIKYFNMYIDGLPIPLPYVKKKGLDQVPELGEVEIRKAFADSEMTEAVND 213
OY 134 -----ELKMAPIETFFY-----HEGDFEGKGEITMSYA--DEBEARVITMK 174
DB 214 WPIPESEVPDYKRAVIDIEVYTPKGRIPDEKAEF---PITISLAGNOSTKRVVL 270
OY 175 NIDL-----PYVDVSNEREMIKRFOVJIREKDPVLITYNGDNPLPYLIRAKEL 226
DB 271 REDVNSQITKHVDIYETKSERELIRREFDIT--LDVPIITFNDDDPDIYVYRALKL 328
OY 227 GVTLLGRDKHEPEPKIRHMGDSFAVEIKGRIFHDLFP--VARRT-----NLPTYLE 278
DB 329 NPT-----PELIPFDIINDEGKYLA-----GIHIDLYKFFFNRAIRNAYFEKQYNEVND 378
OY 279 AVEVAVLKTSKSLAEELTAIETESMKKLAQYSMEDARATYEL---GKEFPMEAE 334
DB 379 AVATLIDGMSKYL--DTLISFLD---LDKLIENSRDAELTKLTFFNNNLVKKLITL 432
OY 335 IAKLIGOSVMDVSRSTG---NLVEMYLLEAVYERNEIAPNKPDEEYRRRLRTT--- 386
DB 433 LARISKMGLEELITREVSFWINKLYWEHRR---RNMILIPKEELIRSSQIKTAALIK 488
OY 387 ---YLGCVYKPERKGLMENTYLDPRCLPSIITVHNVSPTLREGECKN---YDAPI 439
DB 489 GKRYGAVVIVPPAGVFNVNVLDPASLYSIIRMNMSIYEVVDENCKNKEYVDETGE 548
OY 440 VGYKFCQDFGFIPTSIGELITMRQEI--KKKKA--TIPIEKMMIDYQRAVKLANSTY 497
DB 549 VLIHTCKDKPGITAVITGLIRDFRKYVYKKAKKSONISEQGSYDVQRAKAVINMTY 608
OY 498 GYMGYPAKRWYSKECAESVTANGRYIEMTIKEIEKEFGKVLVADTDFGYATIGEKPE 557
DB 609 GYFGAENFPLVAPAAVASVTAIGRVITTYVYC--RSJGLQVLYDGTDSMFLMNS---- 663
OY 558 TIKKAKAEFLKYNSKLGLGLELEYEGFLRGFPVA---KKRVAVIDEGGITTIGLEV 613
DB 664 --KEKLEELIKFVKGKFG--LDLEVDKIYK---FVAFSGLKNNYGVYPDGCTIDJGMLA 716
OY 614 VRDMSLEIAKETQAKVLEAILKEDSVKAVEI-----VKDVEIKKYQVPLEKLVTH 666
DB 717 KKRNTPEFLKKEFNEVKQVLTITNSPDPIKIRDOLEKIKIKIYKLNKNGYNDELFLR 776
OY 667 EQITKDLSEYKAIGP--HVAIAKRLAKGIRKVPCTIISYIVLRSGKISD--RVILLSEYD 724

```

Db 777 VMSKPLESTKNTPOHVKALOLRSGVAVLPDRIIMVKKSGDKGVOLAKUSE-- 834
QY 725 PKRHKYDPDYIENOVLPVLRILEAFG 752
Db 835 -----IDVDKYID-AVNSTFEQILKAFG 856

RESULT 14
Q91VT0
ID Q91VT0 PRELIMINARY; PRT; 1105 AA.
AC Q91VT0;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE DNA POLYMERASE DELTA 1, CATALYTIC DOMAIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST TUMOR;
RA Strausberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009128; AAH09128.1; -
SQ SEQUENCE 1105 AA; 123744 MW; 20323690DDA72406 CRC64;

Query Match 15.3%; Score 616.5; DB 11; Length 1105;
Best Local Similarity 27.2%; Pred No. 6.6e-28;
Matches 206; Conservative 111; Mismatches 297; Indels 123; Gaps 27;

QY 60 GKIVRVDAVKKKFLGRDVEY--WKLFENP-----QDVPLRGKIREHPAVIDIYE 111
Db 269 GKIVR-----RAKKATLQLEVDVLSDVISHPEGQORAPLR-----VLS 312
QY 112 YDIPFAKRYLIDKGLIP-MEGDEELKIMAFDIETFYHEGDEFGKGIIMISYADEEARV 170
Db 313 FDEECAGR-----GIFPERPDVIOICSLGR--WGEPEPLRLALTLRPCAPILGAKY 366
QY 171 ITWKNDLPYDVVNSEREMIKRFOIVREKDPDLITYGDNFDDPYLIRAEKLGVL 230
Db 367 QSE-----KEEDLQAMADFLAMDDBDVTGYNQFNFDLYLSRAQILKVDK 415
QY 231 --LLG-----RKENHEPKIRMGDSFAVEIKGRIFHDLFPVYRRTINLPYTLA 279
Db 416 FPELGEVGTGLRSNIRDSFOSRQVGR--DSKVISMVGRQMDLQVLLREHKLRSTLNA 474
QY 280 YVEAVLGKTKSKLGADEIAIWE--TEESMKKLAQYSMEDARATYELGKEFFPM----- 331
Db 475 VSEHFLGEQKEDVQHSITTDLQNGNEQTRRLRAVYCLKDA-----FLPLRLLERLM 525
QY 332 ---EAEKLALIGOSV-MWVSRSSTGNLVEWYLLRVAVERNELAPNKPDE--EYERRRLR 384
Db 526 VLVNNVEMARVTVGLYLTLRGQYKVS--QLLRQAMRGGLMPVYKREGSD----- 578
QY 385 TTYLGGVKEPEEGLME-NITYLDFRCLYPSIIIVTHNSPDLERBGC-----KNYDV 436
Db 579 --VTGATVIEPLKGYVDVPLATLDFESSLYPSIMAHNLCTTLTLRGAQAQKLGKDEF 636
QY 437 APIVGRKFCDF--PGFISILGELTMRQELKKMKATIDPIEKKMLDYRQAVLHNN 494
Db 637 KPTGGEFVAKSSVRKGLPOLLENLISARRAKAVELAQETDPLRQVLDGRQLALVSN 696
QY 495 SYGYGMYPKARWYSKECAESVTAMGRHYIEMTIKEIEKF-----GFLVYADTDC 546
Db 697 SYVGFGAOGVKLPCLFIEQSOTVGFGRMTIEKTKOLVESYTYENGYDANAKVYGDTS 756
QY 547 EVATIGEPETIKKAKELKYNKLPOLLELEVEGYLGRFVAKKRYAVI----- 600
Db 757 VMCRFVSSVAEAMSLGREGAANVSSHPSPIRLEFEKVFYPLLISKRYAGLFRSSS 816
QY 601 DEGRITTTGLEVVRDMSIAKETQAKVLEALIKEDSEKAVEIKVDYVEETIAKIQVPL 660

Db 817 DAHDKMDCKGLEAVRRDNCPLVANLYTSSLRRLIVRDPGGAHAHAKDYISDLGNRIDI 876
QY 661 EKIVHIOITKDLSEYKAIGPVAIAKRLAKGIVRP--GTIISYIVRGSG-----K 712
Db 877 SOLVITKELTRAADYAGKOAHELAEKRRKPPGSAFSLGDNVPYIIGAAKGAVALYK 936

QY 713 ISDRVILLSEYDPKHKYDPDYIENOVLPVLRILE 749
Db 937 SEDPLFVLEHSLP-----IDTQYVLEQOLAKPLRLIFE 969

RESULT 15
Q96H98
ID Q96H98 PRELIMINARY; PRT; 1107 AA.
AC Q96H98;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DE POLYMERASE (DNA DIRECTED), DELTA 1, CATALYTIC SUBUNIT
DE (125KD).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE-LYMPH, AND LYMPHOMA;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008800; AAH08800.1; -
SQ SEQUENCE 1107 AA; 123611 MW; 10A8A3271916EDBA CRC64;

Query Match 15.3%; Score 616; DB 4; Length 1107;
Best Local Similarity 27.5%; Pred No. 7.1e-28;
Matches 202; Conservative 134; Mismatches 309; Indels 90; Gaps 24;

QY 68 AVKVKRKLGRDVE--VWKLIFENPDVPLRGKIREHPAVIDIYEDIPFAKRYLIDK 124
Db 274 ALRKREKATQCLEADVLSMDVSHPEGPWQR-----IAPLRVLSFDEECAGR-----K 323
QY 125 GLIP-MEGDEELKIMAFDIETFYHEGDEFGKGIIMISYADEEARVITWKNDLPYDV 183
Db 324 GIFPERPDVIOICSLGR--WGEPEPLRLALTLRPCAPILGAKVQSE----- 372
QY 184 VNSEREMIKRFOIVREKDPDLITYGDNFDDPYLIRAEKLGVL--LLG----- 233
Db 373 --KEEDLQAMSFITRIMPDPVTGYNQFNFDLYLSRAQILKVDKTPFPLRGVAGLCSN 430
QY 234 -RKENHEPKIRMGDSFAVEIKGRIFHDLFPVYRRTINLPYTLAAYVEAVLGKTKSKL 292
Db 431 IRDSSFOSQOTGR--DTKVISMVGRQMDLQVLLREYKLRSTLNAVSFHFLEGEQEDV 489
QY 293 GAEFIAIWE--TEESMKKLAQYSMEDARATYELGKEFFPM--EAEKLALIGOSV-MWVS 348
Db 490 QHSITTDLQNGNDQTRRLRAVYCLKDAYLPLRLEELMLVNAVEMARVTVGLSYLSR 549
QY 349 SSTGNLVEWYLLRVAVERNELAPNKPDE--EYERRRLRTYTLGGVKEPEEGLME-NITY 405
Db 550 GQGVKVS--QLLRQAMHGLMPRVKSEGED-----YTGATVIEPLKGYVDVPIAT 600
QY 406 LDFCLYPSIIIVTHNSPDLERBGC-----CKNYDVAIVGKFCDF--PGFISIL 456
Db 601 LDFSSLYPSIMAHNLCTTLRPGTAQKLGLEDQFITPFGDEVKTSVRKGLLPOL 660
QY 457 GELITTRQELKKMKATIDPIEKKMLDYRQAVLHNSYGYGMYPKARWYSKECAESV 516
Db 661 ENLSARKRAKAELETDPRLRQVLDGRQLALKVSANVSYFTGQVQKLPCLFIEQSOSV 720
QY 517 TMAGRIYIMTJKEIEKF-----GFLVYADTDCGYAIIIPGKRPETIKKAKELFK 568
Db 721 TGFGRQMIKTKOLVESKTYENGISTAKVYVGDYVMCRFVSSVAEAMALGREAD 780

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2002, 14:18:07 ; Search time 14.97 Seconds
(without alignments)
2001.933 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026
Sequence: 1 MIFPDYITKDGKPIIRPK.....KEDLKYSKQVGLDAWLKK 774

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	3446.5	85.6	1829	DPOL_THEST
2	3356	83.4	1702	DPOL_THELI
3	3253.5	80.8	773	DPOL_THEGO
4	3204.5	79.6	775	DPOL_THEGO
5	3164.5	78.6	771	DPOL_PYRAB
6	3163.5	78.6	771	DPOL_PYRAB
7	3151	78.3	775	DPOL_PYRAB
8	2974.5	72.9	1312	DPOL_PYRAB
9	2886	71.7	1235	DPOL_PYRAB
10	2809	69.8	1523	DPOL_PYRAB
11	2782	69.1	1671	DPOL_PYRAB
12	2769.5	68.8	1699	DPOL_PYRAB
13	2535.5	63.0	1666	DPOL_PYRAB
14	1392	34.6	824	DPOL_MERVO
15	1301	32.3	781	DPOL_MERVO
16	1292	32.1	1634	DPOL_MERVA
17	1209.5	30.0	586	DPOL_MERVA
18	1138	28.3	784	DPOL_MERVA
19	697.5	17.3	1105	DPOL_MERVA
20	693.5	17.2	1088	DPOL_MERVA
21	691	17.2	763	DPOL_MERVA
22	676.5	16.8	764	DPOL_MERVA
23	676.5	16.8	1081	DPOL_MERVA
24	657.5	16.3	1038	DPOL_MERVA
25	657.5	16.3	1086	DPOL_MERVA
26	624.5	15.5	1092	DPOL_MERVA
27	621.5	15.4	1097	DPOL_MERVA
28	619.5	15.4	1103	DPOL_MERVA
29	616.5	15.3	872	DPOL_MERVA
30	616.5	15.3	882	DPOL_MERVA
31	615.5	15.3	1105	DPOL_MERVA
32	615.5	15.3	1103	DPOL_MERVA
33	614.5	15.3	1094	DPOL_MERVA

ALIGNMENTS

RESULT ID	DPOL_THEST	STANDARD	PRT: 1829 AA.
34	611	15.2	1107
35	609	15.1	1513
36	605.5	15.0	1106
37	587	14.6	1492
38	584	14.5	1451
39	583.5	14.5	1451
40	581	14.4	1081
41	578.5	14.4	875
42	570.5	14.2	959
43	570	14.2	1465
44	569.5	14.1	1243
45	548.5	13.6	1339

DPOL_THEST 1
ID DPOL_THEST STANDARD: PRT: 1829 AA.
AC 033845:
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL.
OS Thermococcus sp. (strain TY).
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID=110163;
RN [1]
RX MEDLINE: 88994267; PubMed: 9434178;
RA Methanospirillum H. Frey B. Methanospirillum G.;
RT from the hyperthermophilic archaeon Thermococcus sp. Ty. 1;
RL Gene 204:153-158 (1997).
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate + [DNA](N).
CC -1- PPM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES A POST-TRANSLATIONAL EXCISION OF THE THREE INTERVENING REGION (INTERNS) FOLLOWED BY PEPTIDE LIGATION.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: Y13030; CAA73475.1; -.
CC HSP: P56689; JTCO.
DR InterPro: IPR002064; DNA_POL_B.
DR InterPro: IPR003586; HINTC.
DR InterPro: IPR003587; HINTC.
DR InterPro: IPR004203; Intein.
DR InterPro: IPR004042; Intein_endonuc.
DR Pfam: PF00136; DNA_POL_B; 4.
DR Pfam: PF03104; DNA_POL_B_exo; 1.
DR PRINTS: PR00379; INTEIN.
DR SMART: SM00305; HINTC; 3.
DR SMART: SM00306; HINTC; 3.
DR SMART: SM00486; POLB; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
DR PROSITE: PS00881; PROTEIN_SPLICING; 3.
KW TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolyase; Endonuclease; Autocatalytic cleavage;
KW Protein splicing.
FT CHAIN 1 409 DNA POLYMERASE, 1ST PART (POTENTIAL).
FT CHAIN 410 769 INTEIN I.
FT CHAIN 770 855 DNA POLYMERASE, 2ND PART (POTENTIAL).
FT CHAIN 856 1392 INTEIN II.

FT CHAIN 1393 1441 DNA POLYMERASE, 3RD PART (POTENTIAL).
 FT CHAIN 1442 1598 INTEIN II.
 FT CHAIN 1599 1829 DNA POLYMERASE, 4TH PART (POTENTIAL).
 SO SEQUENCE 1829 AA; 211875 MW; A113AB0C57EB9CB3 CRC64;

Query Match 85.6%; Score 3446.5; DB 1; Length 1829;
 Best Local Similarity 42.2%; Pred. No. 3.2e-186;
 Matches 771; Conservative 0; Mismatches 3; Indels 1055; Gaps 3;

QY 1 MFDITDTITKDGKPIIRIFKKEGKEFKIELDPHOPITALLDSDAIDEIKAIKGERHG 60
 DB 1 MLDITDTITKDGKPIIRIFKKEGKEFKIELDPHOPITALLDSDAIDEIKAIKGERHG 60
 QY 61 KIVRVDAVAVKKKKFLGRDVEVKLIFEHPDYPALRGKIREHPAVIDIYEYDIPKARY 120
 DB 61 KIVRVDAVAVKKKKFLGRDVEVKLIFEHPDYPALRGKIREHPAVIDIYEYDIPKARY 120
 QY 121 LIDKGLIPMEGDELKLMADITETFEHDEFGKGEIIMISYADEEBARVITWKNDLPY 180
 DB 121 LIDKGLIPMEGDELKLMADITETFEHDEFGKGEIIMISYADEEBARVITWKNDLPY 180
 QY 181 VDVVSNREMIKRFVOIVREKDPDLITTYGNDNFDPYLTKRAEKLGVTLGRDKHPE 240
 DB 181 VDVVSNREMIKRFVOIVREKDPDLITTYGNDNFDPYLTKRAEKLGVTLGRDKHPE 240
 QY 241 PKIHRGDSFAVEIKGRHEDLPVVRRTINLPTYLEAVYEAVALGKTSKLGAEETAAI 300
 DB 241 PKIHRGDSFAVEIKGRHEDLPVVRRTINLPTYLEAVYEAVALGKTSKLGAEETAAI 300
 QY 301 WEFEESKKLAQYSMEDARTYELGKEFFPMEAEKALIGOSWDVSRSTGULVEMYLL 360
 DB 301 WEFEESKKLAQYSMEDARTYELGKEFFPMEAEKALIGOSWDVSRSTGULVEMYLL 360
 QY 361 RYVERNELAPNKPDEEYRRLRTTYLGKYKPERGKLENTIYDIFRC----- 410
 DB 361 RYVERNELAPNKPDEEYRRLRTTYLGKYKPERGKLENTIYDIFRC----- 410
 QY 411 ----- 410
 DB 421 GKGIWISDVKEGDYILIGDMQVKKWYHGEKLININGLKCPNKHVPVVTENDRQ 480
 QY 411 ----- 410
 DB 481 TRIRDSLANSFLSGKVGKIIITTKLEKIAEFENKPNPSEELIKGSLIILAEGLTLRK 540
 QY 411 ----- 410
 DB 541 DIEFDSSRCKKRISHQYREITIGENEKELLERILYIPDKLFGIRSVAKKGDJNAKTI 600
 QY 411 ----- 410
 DB 601 TPAKKAVALQIEELLKNIESLAPAVLRFGEFBDATVNRIRSTIVTGTGNMKKIDIVA 660
 QY 411 ----- 410
 DB 661 KLLDSGIPYSRYEYKYIENGKELTKHILETTGRDGLILFQTLVGFISSKNEALEKAIE 720
 QY 411 ----- 410
 DB 721 VREMNRLKNNSTFEVYSSEYKGEVYDLLEGNPYFANGILTHNLSYISITVTHN 780
 QY 421 VSPDTLERECCKNYDAPIVGKCKDPGFIPIISIGELITMROEIKKKKATIDPIEKK 480
 DB 781 VSPDTLERECCKNYDAPIVGKCKDPGFIPIISIGELITMROEIKKKKATIDPIEKK 480
 QY 481 MLYRQRAVLAH----- 493
 DB 841 MLYRQRAVLAH----- 493
 QY 494 ----- 493
 DB 901 VDNIFAFSLNKESKSEIKVKALLIRHKYKGEAYEVLANSGRKIHITRGHSLFTIRNGKI 960

QY 494 ----- 493
 DB 961 KEIMGEVYKVGDLIIVPKKYLKNEKAVINIPBLISKLPEDOTADYVMTTPVGRKNQFK 1020
 QY 494 ----- 493
 DB 1021 GMLRTLKWIKEESKRIRTEPNRYLFLHEELGFVKLLPRGYEYTDWEGKLRRLQYELKVK 1080
 QY 494 ----- 493
 DB 1081 NLRNGNKKREYLVRENDIKDSVSCFPKKELEEMKIGTXGFRKCKILKYDEDFGKFLGY 1140
 QY 494 ----- 493
 DB 1141 VSEGYAGAOKNKTGMSYSVKLYNENPNVLKDKMKNIAEKFFGKVRVGNKNCVIDPKKMYL 1200
 QY 494 ----- 493
 DB 1201 LAKSLGVTAENKRIPSIIFDSSEPYRMFLRAYFYGGDIDHPSKRLRSTKSELLANOL 1260
 QY 494 ----- 493
 DB 1261 VELLNSLGVSSTIGFDSGYRYVINEDLPLOTSRQKNTYNNLIPKYLEETFGRRKQ 1320
 QY 494 ----- 493
 DB 1321 KNITFEKRELADSGKLDKRRVKLDFLNGDIVLRVKNVKEKREYGYVYDLSVEDNEN 1380
 QY 494 ----- 542
 DB 1381 FLVFGLLVAHNSYGYGMGPARKWYSKECAESVTAMGRHYIEMTJKEIEKGFVKVLA 1440
 QY 543 D----- 543
 DB 1441 DSVTGTETIYKRNKGRLEFPPIKLEPRVDYRIGEKYCLIEDVEALITLDNRKILMKKY 1500
 QY 544 ----- 543
 DB 1501 PYVRRRAKKKYYRIMTNSWYIDVTEHSLIIVAEDELKARPMEIEGKSLIATKDDLSG 1560
 QY 544 ----- 543
 DB 1561 VEXIKPHAIEISYNGVYDIEVEGTHRFANGILVHNTDGFATIPGEKPEIIRKRAK 1620
 QY 566 FLKYINSKLPGLLELEVEGEYLRGFFVAKKRYAVIDEGRITRGLEVVARDSEIAKET 625
 DB 1621 FLKYINSKLPGLLELEVEGEYLRGFFVAKKRYAVIDEGRITRGLEVVARDSEIAKET 625
 QY 626 QAKVLEAILLKEDSVEKAVEIVKQVVEIAKYOVLKELVTHEQITDLSYKAIQPHVAI 685
 DB 1681 QAKVLEAILLKEDSVEKAVEIVKQVVEIAKYOVLKELVTHEQITDLSYKAIQPHVAI 685
 QY 686 AKRLAKGKIVRGCTIISYVLRSGKISDRVILLSEYDPKKKKYPODYIENOVPAVL 745
 DB 1741 AKRLAKGKIVRGCTIISYVLRSGKISDRVILLSEYDPKKKKYPODYIENOVPAVL 745
 QY 746 RILEAFGRKEDLKYOSSKQVGLDAMLK 774
 DB 1801 RILEAFGRKEDLKYOSSKQVGLDAMLK 774
 QY 1801 RILEAFGRKEDLKYOSSKQVGLDAMLK 1829

RESULT 2
 DPOL THELI STANDARD; PRT: 1702 AA.
 AC P30317:
 DT 01-APR-1993 (rel. 25, Last sequence update)
 DT 01-APR-1993 (rel. 25, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7) (Vent DNA polymerase) [Contains:
 DE Endonuclease P1-T111 (EC 3.1.-.-) (T11 pol-1 intein) (IVPS2);
 DE Endonuclease P1-T111 (EC 3.1.-.-) (T11 pol-2 intein) (IVPS1)].
 GN POL.

```

0S Thermococcus litoralis.
0C Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
0X NCBI_taxonomy:2265;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE:92302285; PubMed:1608969;
RA Perler F.B., Combs D.G., Jack W.E., Moran L.S., Qiang B.,
RA Kucera R.B., Benner J., Slatko B.E., Nwankwo D.O., Hempstead S.K.,
RA Carlow C.K.S., Jannasch H.;
RT "Intervening sequences in an Archaea DNA polymerase gene.;"
RL Proc. Natl. Acad. Sci. U.S.A. 89:5577-5581(1992).
RN [2]
RP PROTEIN SPICING.
RX MEDLINE:93117083; PubMed:1475179;
RA Hodges R.A., Perler F.B., Noren C.J., Jack W.E.;
RT "Protein splicing removes intervening sequences in an archaea DNA
RT polymerase.;"
RL Nucleic Acids Res. 20:6153-6157(1992).
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC -1- FUNCTION: IN THE INTEIN SECTION, BELONGS TO THE HOMING
CC ENDOUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE
CC INTEIN.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + {DNA}(N).
CC -1- PMW: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION
CC (INTEREINS) FOLLOWED BY PEPTIDE LIGATION.
CC -1- BIOTECHNOLOGY: used in the PCR method because of its high
CC thermostability and low error rate. Sold by New England Biolabs.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
CC ENDOUCLEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M74198; AAA72100.1; -.
DR EMBL; M74198; AAA72101.1; -.
DR PIR; S42459; S42459.
DR HSSP; P56689; ITCO.
DR REBASE; 2613; PI-T111.
DR REBASE; 2621; PI-T1111.
DR InterPro; IPR002064; DNA_pol_B.
DR InterPro; IPR003587; HincC.
DR InterPro; IPR003587; HincC.
DR InterPro; IPR002203; Intein.
DR InterPro; IPR004042; Intein_endonuc.
DR Pfam; PF00136; DNA_pol_B; 3.
DR Pfam; PF03104; DNA_pol_B_exo; 1.
DR PRINTS; PR00379; INTEIN.
DR SMART; SM00305; HincC; 2.
DR SMART; SM00306; HincC; 2.
DR SMART; SM00486; POLEB; 1.
DR PROSITE; PS00116; DNA_POLYMERASE_B; FALSE_NEG.
DR PROSITE; PS00881; PROTEIN_SPLICING; 2.
KW Transase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
KW Protein splicing; Intron homing.
FT CHAIN 1 494 DNA POLYMERASE, 1ST PART.
FT CHAIN 495 1032 ENDOUCLEASE PI-T11 II.
FT CHAIN 1033 1081 DNA POLYMERASE, 2ND PART.
FT CHAIN 1082 1471 ENDOUCLEASE PI-T11 I.
FT CHAIN 1472 1702 DNA POLYMERASE, 3RD PART.
SQ SEQUENCE 1702 AA; 197293 MW; 21D6B98C75F53B20 CRC64;

```

Query Match	83.4%	Score 3356;	DB 1;	Length 1702;
Best Local Similarity	42.7%;	Pred. No.3,6e-181;		
Matches 726;	Conservative	33;	Mismatches 15;	Indels 928;
				Gaps 2;
QY 1 MFEDDYITKDKPPIIRIFPKKENGGEFKIEDLDPHOPYIALLKDDSAIDIEIKAIINGERG 60	1 MFEDDYITKDKPPIIRIFPKKENGGEFKIEDLDPHOPYIALLKDDSAIDIEIKAIINGERG 60			
Db 1 MLDDYITKDKPPIIRIFPKKENGGEFKIEDLDPHOPYIALLKDDSAIDIEIKAIINGERG 60	1 MLDDYITKDKPPIIRIFPKKENGGEFKIEDLDPHOPYIALLKDDSAIDIEIKAIINGERG 60			
QY 61 KIVRVDAVKYKKFELGRDVEVWKLFIEHPDVPALRCRIKREHPAVIDIEYDIFPAKRY 120	61 KIVRVDAVKYKKFELGRDVEVWKLFIEHPDVPALRCRIKREHPAVIDIEYDIFPAKRY 120			
Db 61 KIVRVDAVKYKKFELGRDVEVWKLFIEHPDVPALRCRIKREHPAVIDIEYDIFPAKRY 120	61 KIVRVDAVKYKKFELGRDVEVWKLFIEHPDVPALRCRIKREHPAVIDIEYDIFPAKRY 120			
QY 121 LIDKLLIPGEGEELKLAFOIETFYHDEDEGKEEIMISVADDEEAVITWKNIDLPY 180	121 LIDKLLIPGEGEELKLAFOIETFYHDEDEGKEEIMISVADDEEAVITWKNIDLPY 180			
Db 121 LIDKLLIPGEGEELKLAFOIETFYHDEDEGKEEIMISVADDEEAVITWKNIDLPY 180	121 LIDKLLIPGEGEELKLAFOIETFYHDEDEGKEEIMISVADDEEAVITWKNIDLPY 180			
QY 181 VDVVNEREMIRFVQIYREKDPDLITYNGNFDLPYLIRKRAEKLGVYLLGRDKHEPE 240	181 VDVVNEREMIRFVQIYREKDPDLITYNGNFDLPYLIRKRAEKLGVYLLGRDKHEPE 240			
Db 181 VDVVNEREMIRFVQIYREKDPDLITYNGNFDLPYLIRKRAEKLGVYLLGRDKHEPE 240	181 VDVVNEREMIRFVQIYREKDPDLITYNGNFDLPYLIRKRAEKLGVYLLGRDKHEPE 240			
QY 241 PKIHMGDSFAVEIKGRIHEDLPVAVRRTINLPYTLAEVAVYLGKTSKIGAEIATAI 300	241 PKIHMGDSFAVEIKGRIHEDLPVAVRRTINLPYTLAEVAVYLGKTSKIGAEIATAI 300			
Db 241 PKIHMGDSFAVEIKGRIHEDLPVAVRRTINLPYTLAEVAVYLGKTSKIGAEIATAI 300	241 PKIHMGDSFAVEIKGRIHEDLPVAVRRTINLPYTLAEVAVYLGKTSKIGAEIATAI 300			
QY 301 WTEESMKKLAQYSEDARATYELGKEFPMAEELAKLIGOSVMPVSRSTGNLVEWYLL 360	301 WTEESMKKLAQYSEDARATYELGKEFPMAEELAKLIGOSVMPVSRSTGNLVEWYLL 360			
Db 301 WTEESMKKLAQYSEDARATYELGKEFPMAEELAKLIGOSVMPVSRSTGNLVEWYLL 360	301 WTEESMKKLAQYSEDARATYELGKEFPMAEELAKLIGOSVMPVSRSTGNLVEWYLL 360			
QY 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKEBERGLMENTYLDPRCLYPSIITVTHN 420	361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKEBERGLMENTYLDPRCLYPSIITVTHN 420			
Db 361 RYAVARNELAPKPDDEEYRRRLRTTYLGGYKEBERGLMENTYLDPRCLYPSIITVTHN 420	361 RYAVARNELAPKPDDEEYRRRLRTTYLGGYKEBERGLMENTYLDPRCLYPSIITVTHN 420			
QY 421 VSPDLEPRGGCNVAVAPVGVYKCKDPGCFIPSLIGELITMROELKKMKATIPIEKK 480	421 VSPDLEPRGGCNVAVAPVGVYKCKDPGCFIPSLIGELITMROELKKMKATIPIEKK 480			
Db 421 VSPDLEPRGGCNVAVAPVGVYKCKDPGCFIPSLIGELITMROELKKMKATIPIEKK 480	421 VSPDLEPRGGCNVAVAPVGVYKCKDPGCFIPSLIGELITMROELKKMKATIPIEKK 480			
QY 481 MLDYQRAVVKLTA----- 493	481 MLDYQRAVVKLTA----- 493			
Db 481 MLDYQRAVVKLTA----- 493	481 MLDYQRAVVKLTA----- 493			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 541 VNMLEAFSEFNKKIKSEYKKVAKALIRHKYKGRAYETIOLSSGRRINTAGHSLETVRNGEI 600	541 VNMLEAFSEFNKKIKSEYKKVAKALIRHKYKGRAYETIOLSSGRRINTAGHSLETVRNGEI 600			
Db 541 VNMLEAFSEFNKKIKSEYKKVAKALIRHKYKGRAYETIOLSSGRRINTAGHSLETVRNGEI 600	541 VNMLEAFSEFNKKIKSEYKKVAKALIRHKYKGRAYETIOLSSGRRINTAGHSLETVRNGEI 600			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 601 KEVSGDGIKGGDLIYAPPKIKILNKGVSINIEBLLSDSEBETADIVMIIISAKGRNPFK 660	601 KEVSGDGIKGGDLIYAPPKIKILNKGVSINIEBLLSDSEBETADIVMIIISAKGRNPFK 660			
Db 601 KEVSGDGIKGGDLIYAPPKIKILNKGVSINIEBLLSDSEBETADIVMIIISAKGRNPFK 660	601 KEVSGDGIKGGDLIYAPPKIKILNKGVSINIEBLLSDSEBETADIVMIIISAKGRNPFK 660			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 721 SVKYNKRRKREYLVMENEIKDIPISYPOKELEEMKIGTLNGFPRTNCILKYDEDFGKLGGY 780	721 SVKYNKRRKREYLVMENEIKDIPISYPOKELEEMKIGTLNGFPRTNCILKYDEDFGKLGGY 780			
Db 721 SVKYNKRRKREYLVMENEIKDIPISYPOKELEEMKIGTLNGFPRTNCILKYDEDFGKLGGY 780	721 SVKYNKRRKREYLVMENEIKDIPISYPOKELEEMKIGTLNGFPRTNCILKYDEDFGKLGGY 780			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 781 VSEGTAGAKKNTGTGISYVKLYNDDPNVLEBKMNAVEKFPCKVAVDRNCVSIKKMAYL 840	781 VSEGTAGAKKNTGTGISYVKLYNDDPNVLEBKMNAVEKFPCKVAVDRNCVSIKKMAYL 840			
Db 781 VSEGTAGAKKNTGTGISYVKLYNDDPNVLEBKMNAVEKFPCKVAVDRNCVSIKKMAYL 840	781 VSEGTAGAKKNTGTGISYVKLYNDDPNVLEBKMNAVEKFPCKVAVDRNCVSIKKMAYL 840			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 841 VMKCTGALAEKRIPSVILTSPBEVWMSFLEAYFTGGDIIHPSKRFRLSTSELANOL 900	841 VMKCTGALAEKRIPSVILTSPBEVWMSFLEAYFTGGDIIHPSKRFRLSTSELANOL 900			
Db 841 VMKCTGALAEKRIPSVILTSPBEVWMSFLEAYFTGGDIIHPSKRFRLSTSELANOL 900	841 VMKCTGALAEKRIPSVILTSPBEVWMSFLEAYFTGGDIIHPSKRFRLSTSELANOL 900			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY 901 VFLNLSLGISSVKIGFDSCVRYVYINEDLOFPQTSREKNTYISNLIPEILRDVCGKEQ 960	901 VFLNLSLGISSVKIGFDSCVRYVYINEDLOFPQTSREKNTYISNLIPEILRDVCGKEQ 960			
Db 901 VFLNLSLGISSVKIGFDSCVRYVYINEDLOFPQTSREKNTYISNLIPEILRDVCGKEQ 960	901 VFLNLSLGISSVKIGFDSCVRYVYINEDLOFPQTSREKNTYISNLIPEILRDVCGKEQ 960			
QY 494 ----- 493	494 ----- 493			
Db 494 ----- 493	494 ----- 493			
QY				

```

Db 1021 FLVFGFLYAHNSYGYMGYPKARWYSKECAESVTAMGRHYIMTIREIEKGFVLYA 1080
Qy 543 D----- 543
Db 1081 DVSAGESEIIIRONGKIRPVKIKDLESKVDYSIGKEKCYCLEGEVALLTDDGKLVKPY 1140
Qy 544 ----- 543
Db 1141 PYVMRRANKRMFRILWNSWYIDVTEHSLIGLNTSKTKTAKKIGERLKEVKEPPELGK 1200
Qy 544 ----- 543
Db 1201 AKSLICPNAPLKDENTKTSIAVKFWEVLVLIVGDNGGDSRMAEYVGLSTGKDAEE 1260
Qy 544 ----- 543
Db 1261 IKOKLEPLKTYGVISNYYPKNEKGDNLAKSLVKFMKRFKDEKGRKKIPPEMTELVY 1320
Qy 544 ----- 543
Db 1321 TYIEAFGLRGLFSADGVTTIRKGVPEIRLNIADAFLEVRKLLMIVIGSISIFAETTPNR 1380
Qy 544 ----- 543
Db 1381 YNGVSTGYTKHLRINKKMPAERIGFLTERKOKLLEHLKSAKRVKRNITDEGFDLVHWK 1440
Qy 544 ----- 543
Db 1441 KVEEIPYEGVYDIEVEETHRFANNILVHNTDGFATIPGEKPELKKKAKFLWYINS 1500
Qy 544 ----- 543
Db 1501 KLPGLELELEEGFYLRGFFVAKKRYAVIDEGRITTRGLEVVRRDSEIAKETQAKVLEA 1560
Qy 573 KLPGLELELEEGFYLRGFFVAKKRYAVIDEGRITTRGLEVVRRDSEIAKETQAKVLEA 632
Db 1501 KLPGLELELEEGFYLRGFFVAKKRYAVIDEGRITTRGLEVVRRDSEIAKETQAKVLEA 1560
Qy 633 ILKEDSVKAVEVVKDVEIKAYOYPLKELVTHEOITDLSYKAGISHVATAKRLAK 692
Db 1561 ILKESGVEAVKVEVYKRYEAKRYPLEKLYHEOITDLDKDKYKIGPVVALAKRLAK 1620
Qy 693 GIKVRPGTIIYVLRGSGKISDRVILLSEYDPKHKHYPDYIENQVLPVLRILAEAG 752
Db 1621 GAKVKKGTIIYVLRGSGKISDRVILLSEYDPKHKHYPDYIENQVLPVLRILAEAG 1680
Qy 753 YRKEDLKYSKQVGLDANLKK 774
Db 1681 YRKEDLKYSKQVGLDANLKK 1702

```

RESULT 3

DB: THEGO STANDARD: PRT: 773 AA.

AC P56689;

DT 15-JUL-1999 (Rel. 38, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE DNA polymerase (EC 2.7.7.7) (TO POL).

GN POL OR POLA.

OS Thermococcus gorgonarius.

OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.

OX NCBI_TaxID-71997;

RN (11)

RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).

RX MEDLINE-99199230; PubMed-10097083;

RA Hopfner K.-P., Eichinger A., Engh R.A., Laue F., Ankenbauer W., Huber R., Angerer B.;

RT "Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius."

RL Proc. Natl. Acad. Sci. U.S.A. 96:3600-3605(1999).

CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.

CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate + (DNA)(N).

CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

```

DR PDB: 1TGO; 22-MAR-99.
DR InterPro: IPR002064; DNA_pol_B.
DR Pfam: PF00136; DNA_pol_B_1.
DR Pfam: PF03104; DNA_pol_B_exo_1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBc_1.
DR PROSITE: PS00116; DNA_POLYMERASE_B_1.
DR Transferrase: DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolyase; Nuclease; Exonuclease; Multifunctional enzyme;
KW 3D-structure.
FT DISULFID 428 442
FT DISULFID 506 509
SQ SEQUENCE 773 AA; 89812 MW; F67AF04E875FE844 CRC64;

```

Query Match 80.8%; Score 3253.5; DB 1; Length 773;

Best Local Similarity 78.3%; Pred No. 7.4e-176;

Matches 605; Conservative 83; Mismatches 82; Indels 3; Gaps 2;

```

Qy 1 MIFDDYITKDKSPIIRIKKENGFEKILDPHOPYIYALLKDSALDEIRAKGERRG 60
  1 MIFDDYITKDKSPIIRIKKENGFEKIDYDRNEFYIYALLKDSALIDVKKITAEARRG 60
Db 1 MIFDDYITKDKSPIIRIKKENGFEKIDYDRNEFYIYALLKDSALIDVKKITAEARRG 60
Qy 61 KIVRVYDAVKVKKFLGRDVEWKKLFEHPDYPAKGRIRREHPAYIDYEDIPPAKRY 120
  61 KIVRVYDAVKVKKFLGRDVEWKKLFEHPDYPAKGRIRREHPAYIDYEDIPPAKRY 120
Db 61 KIVRVYDAVKVKKFLGRDVEWKKLFEHPDYPAKGRIRREHPAYIDYEDIPPAKRY 120
Qy 121 LIDKGLIPMEGDEELKMAFDIETFEHGEDEFGKEIIMISYADEEBAHVITWKNDLPY 180
  121 LIDKGLIPMEGDEELKMAFDIETLYHEGEFEFAGSIIMISYADEGCAVITWKNDLPY 180
Db 121 LIDKGLIPMEGDEELKMAFDIETLYHEGEFEFAGSIIMISYADEGCAVITWKNDLPY 180
Qy 181 VDVVSNEREMIKRFVOIVREKDPVLITYNGDNFDLPYLIRAKELIGVTLGLGRDHEP 240
  181 VDVVSNEREMIKRFVIVREKDPVLITYNGDNFDLPYLIRAKELIGVTLGLGRDHEP 240
Db 181 VDVVSNEREMIKRFVIVREKDPVLITYNGDNFDLPYLIRAKELIGVTLGLGRDHEP 240
Qy 241 PKIHRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 300
  241 PKIHRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 300
Db 241 PKIHRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 300
Qy 239 PKIQRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 298
  239 PKIQRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 298
Db 239 PKIQRGDSFAVEIKRIHFEDLPVVRRTINLPYLEVEVAVIGKTSKIGAEIRAI 298
Qy 301 WTEESMKRLAOLYSMEDARATYELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 360
  301 WTEESMKRLAOLYSMEDARATYELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 360
Db 301 WTEESMKRLAOLYSMEDARATYELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 360
Qy 299 WETGEGLEVARVARSMDADVTEELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 358
  299 WETGEGLEVARVARSMDADVTEELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 358
Db 299 WETGEGLEVARVARSMDADVTEELKEFPMEALIKIGOSVMDVSRSTGNLYEWYLL 358
Qy 361 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 420
  361 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 420
Db 361 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 420
Qy 359 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 417
  359 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 417
Db 359 RYAVERNELAPNKPDEEYRRLRTTYLGGYKKEPRLGMENTIYIDFRCLYPSIIVTN 417
Qy 421 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 480
  421 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 480
Db 421 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 480
Qy 418 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 477
  418 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 477
Db 418 VSPDITLREGECKNYDAPIVGYKFCDFPGFIPSLGLDLLEEROKKKMKATIDPIERK 477
Qy 481 MLYROAVAKLANSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIREIEKGFVYL 540
  481 MLYROAVAKLANSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIREIEKGFVYL 540
Db 481 MLYROAVAKLANSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIREIEKGFVYL 540
Qy 478 LLDYORAIKILANSFYGYGAKARWYCKECAESVTAMGRHYIEMTIREIEKGFVYL 537
  478 LLDYORAIKILANSFYGYGAKARWYCKECAESVTAMGRHYIEMTIREIEKGFVYL 537
Db 478 LLDYORAIKILANSFYGYGAKARWYCKECAESVTAMGRHYIEMTIREIEKGFVYL 537
Qy 541 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 600
  541 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 600
Db 541 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 600
Qy 538 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 597
  538 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 597
Db 538 YADTGFYATITGEKPEITKKAKKEFLKINSKLPGLLEBEGFLKRFPAKKRYAVI 597
Qy 601 DEGRITTRGLEVVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 660
  601 DEGRITTRGLEVVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 660
Db 601 DEGRITTRGLEVVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 660
Qy 598 DEBDKTTTGLLEIVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 657
  598 DEBDKTTTGLLEIVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 657
Db 598 DEBDKTTTGLLEIVRRDMSKIKETQAKVLEAILKDSVEKVAEYKDVVEELAKQVPL 657
Qy 661 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 720
  661 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 720
Db 661 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 720
Qy 658 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 717
  658 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 717
Db 658 EKLVIHEOITKDLSEYKAGPVVAIAKRAKGIKVPCTIISYVLRGSGKISDRVILL 717
Qy 721 SEYDPKHKHYDDPYIENQVLPVLRILAEFGYRKEDLKYSKQVGLDANLKK 773
  721 SEYDPKHKHYDDPYIENQVLPVLRILAEFGYRKEDLKYSKQVGLDANLKK 773
Db 721 SEYDPKHKHYDDPYIENQVLPVLRILAEFGYRKEDLKYSKQVGLDANLKK 773

```

RESULT 4

[illegible]

```

OY 361 RAAYERNELAPKPDDEEYRRRLRTTYIGGYKKEEERGLMEITYTLDRCLYPSIITYVN 420
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 359 RRAYERNELAPKPDPERELARR - RGYAGGYKKEEERGLMDNIYVLDPRSLPSIITHN 417
OY 421 VSPDRLERRGCNRYVAPVIVGKCFKDPGFPSPSLIGELITMROEIKKKMATDPIEKK 480
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 VSPDRLNRBCKGEYVAEVEVGHKFKDPGFPSPSLIGDLLEBRQIKKKMATYVPLEKK 477
OY 481 MLDYRQAVKLTANSYGYMGYKARWYKSCAESEVTAMGRHYIMTKEIEKEKGFVYL 540
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 478 LLDYHQRAIKLTANSFYGYGAKARWYCKBAESEVTAMGRYIMVYRLLEERKGFVYL 537
OY 541 YADIDGFPATIDPEKEPEPTIKKAKFEFLYINSKLPGLLEIEYEGFYLNGFVAKKRAVI 600
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 538 YADTDGLHATIDPGADAEFTVKKKKEFLKYINPKLPGLLEIEYEGFYVNGFVTKKYAVI 597
OY 601 DEEGITTRGIEVVRDMSLAKETQAVLEILKEDSEKAVEIVKRVVEILAKYQVPL 660
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 596 DEEGITTRGIEVVRDMSLAKETQAVLEILKEDSEKAVEIVKRVVEILAKYQVPL 657
OY 661 EKLVIHEOTKDLSEYKAGIPHVAITAKRLAAKGIYVRGTTIYSLVSGKISDRIYLL 720
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 658 EKLVIHEOTKDLRLRYKATGPHVAIAAKKLAARGVAKIIRGTYIYSLVLSGSRIGDRALPA 717
OY 721 SEYDPRKKHKYDPPDYIENOVLPVLRILIEAGFYRKEDIKYOSSKOVGLDAMWK 773
    | : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 718 DEFDPTRKHRYDAEYIENOVLPVLRILIEAGFYRKEDIKYOKTKOVGLAMWK 770

RESULT 5
DPOL_PYRAB STANDARD: PRT: 771 AA.
AC P77916;
DT 15-DEC-1998 (Rel. 37, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase 1 (EC 2.7.7.7) (Pab polymerase).
GN POLI OR POL OR PAB1128.
OS Pyrococcus abyssi.
OS Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ST 855;
RA Cammon M., Querellou J., Barbier G., Dietrich J., Forterre P.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAY;
RA Helling R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
    structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
    + [DNA](n).
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: Z54174; CAA90888.1; -
DR EMBL: AJ248288; CAB50625.1; -
DR HSSP: P56689; ITGO.
DR InterPro: IPR002064; DNA_POL-B.
DR Pfam: PF00136; DNA_POL_B.1.
DR Pfam: PF03104; DNA_POL_B_exo; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLB; 1.

```

DR PROSITE: P500116; DNA-POLYMERASE_B.1.
 KW Transferrase; DNA-directed DNA polymerase; DNA replication;
 KM DNA-binding; Complete proteome.
 FT CONFLICT 263 263 V -> A (IN REF. 1).
 FT CONFLICT 267 267 A -> T (IN REF. 1).
 FT CONFLICT 281 281 A -> V (IN REF. 1).
 FT CONFLICT 320 320 F -> S (IN REF. 1).
 FT CONFLICT 339 339 Q -> H (IN REF. 1).
 FT CONFLICT 359 359 R -> T (IN REF. 1).
 FT CONFLICT 391 391 K -> N (IN REF. 1).
 FT CONFLICT 756 756 K -> R (IN REF. 1).
 SQ SEQUENCE 771 AA: 89496 MW: 1104870458A5522 CRC64:

Query Match 78.6%; Score 3164.5; DB 1; Length 771;
 Best local Similarity 76.1%; Pred. No. 7.5e-171;
 Matches 588; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

QY 1 MFDVTYITKDGPIIRIFKKEGKELIDPHFQPIYALDKDSADIEKAIKGRHG 60
 DB 1 MIDADYITEDGPIIRIFKKEGKELIDPHFQPIYALDKDSADIEKAIKGRHG 60
 QY KIYRVDAVAVKKKFLGRPLEVKKLYLEHPQDPALRGKIREHPAVDIYEDIPFAKRY 120
 DB 61 KIYRVDAVAVKKKFLGRPLEVKKLYLEHPQDPALRGKIREHPAVDIYEDIPFAKRY 120
 QY 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 DB 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 QY 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 DB 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 QY 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 DB 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 DB 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 QY 241 PKIHRMDSFAVELKGIHDLFPVVRTINLTPTYLEAVYEAVLCKTSKLGAEITAI 300
 DB 241 PKIHRMDSFAVELKGIHDLFPVVRTINLTPTYLEAVYEAVLCKTSKLGAEITAI 300
 DB 239 PKIHRMDSFAVELKGIHDLFPVVRTINLTPTYLEAVYEAVLCKTSKLGAEITAI 298
 QY 301 WETEESKKLAQYSMEDARATYELGKEFFPMEALAKIGOSVWDVSRSTGNLVEMYL 360
 DB 299 WETGKGLERAVAKYSMEDAKYTFELGKEFFPMEALAKIGOSVWDVSRSTGNLVEMYL 358
 QY 361 RYAVERNELAPNKPDEEYRRRLRTYLLGYVKEPERGMENTIYDFRCLYSIITVHN 420
 DB 359 RYAVERNELAPNKPDEEYRRRLRTYLLGYVKEPERGMENTIYDFRCLYSIITVHN 418
 QY 421 VSPDTEREGCKNVDAPIVGYKFCDFPGFISILGELITMROEIKKKKATIDPIEK 480
 DB 419 VSPDTEREGCKNVDAPIVGYKFCDFPGFISILGELITMROEIKKKKATIDPIEK 478
 QY 481 MLDYRORAVLHANSYGYGYGPRAMYSKCAESVTAMGRHYIEMTKIEEKGFKVL 540
 DB 479 MLDYRORAVLHANSYGYGYGPRAMYSKCAESVTAMGRHYIEMTKIEEKGFKVL 537
 QY 541 YADTDGVAITIPGKPEPTIKKAKKELKYNSKLPGLLELYEGSFYRGEFVAKKRAVI 600
 DB 538 YADTDGVAITIPGKPEPTIKKAKKELKYNSKLPGLLELYEGSFYRGEFVAKKRAVI 597
 QY 601 DEGRATITRLEAVRRMSIAKETQAKVLEALIKEDSVKAVEIVADVVEETAKIYOVL 660
 DB 598 DEGRATITRLEAVRRMSIAKETQAKVLEALIKEDSVKAVEIVADVVEETAKIYOVL 657
 QY 661 EKLIVHBOITKDLSEYKAIPHVAIAKLAAGIKVAPGTIISYIVLRSKGISDRVYL 720
 DB 658 EKLIVHBOITKDLSEYKAIPHVAIAKLAAGIKVAPGTIISYIVLRSKGISDRVYL 717
 QY 721 SEVDPRKKHAKDPDYIYNOVLPAVLRLLEAFGRKEDLKQSSQOVLDAWLK 773
 DB 718 SEVDPRKKHAKDPDYIYNOVLPAVLRLLEAFGRKEDLKQSSQOVLDAWLK 770

ID DPOL_PYRSE STANDARD; PRT; 771 AA.
 AC P77932;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE DNA polymerase (EC 2.7.7.7).
 GN POL.
 OS Pyrococcus sp. (strain GE23).
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_Taxid=86036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Camdon M., Querellon J., Bouyoub A., Raouene G., Barbier G.,
 RA Forterre P., Dietrich J.;
 RL Submitted (SEP-1995) to the EMBL/Genbank/DBA databases.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
 CC + (DNA)(N).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 254173; CAA90887.1; -
 CC HSSP: P56689; ITGO.
 DR InterPro: IPR002064; DNA_POL_B.
 DR Pfam: PF00136; DNA_POL_B.1.
 DR Pfam: PF03104; DNA_POL_B-exo; 1.
 DR PRINTS: PR00106; DNAPOLB.
 DR SMART: SM00486; POLBc; 1.
 DR PROSITE: P500116; DNA-POLYMERASE_B.1.
 KW Transferrase; DNA-directed DNA polymerase; DNA replication;
 KM DNA-binding; 771 AA: 89400 MW: 1506756996056C8 CRC64;
 SQ SEQUENCE

Query Match 78.6%; Score 3163.5; DB 1; Length 771;
 Best local Similarity 75.9%; Pred. No. 8.5e-171;
 Matches 587; Conservative 89; Mismatches 94; Indels 3; Gaps 2;

QY 1 MFDVTYITKDGPIIRIFKKEGKELIDPHFQPIYALDKDSADIEKAIKGRHG 60
 DB 1 MIDADYITEDGPIIRIFKKEGKELIDPHFQPIYALDKDSADIEKAIKGRHG 60
 QY KIYRVDAVAVKKKFLGRPLEVKKLYLEHPQDPALRGKIREHPAVDIYEDIPFAKRY 120
 DB 61 KIYRVDAVAVKKKFLGRPLEVKKLYLEHPQDPALRGKIREHPAVDIYEDIPFAKRY 120
 QY 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 DB 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 QY 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 DB 121 LIDKGLIPMGDELKLMADIEFYHGEDEFGKELIMISYADEEARVITKNDLPY 180
 QY 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 DB 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 DB 181 VDVVSNREMIKRPVQIVREKDPDLITVNGDNFDPYLKRAKAGVTLGDRKHPE 240
 QY 241 PKIHRMDSFAVELKGIHDLFPVVRTINLTPTYLEAVYEAVLCKTSKLGAEITAI 300
 DB 239 PKIHRMDSFAVELKGIHDLFPVVRTINLTPTYLEAVYEAVLCKTSKLGAEITAI 298
 QY 301 WETEESKKLAQYSMEDARATYELGKEFFPMEALAKIGOSVWDVSRSTGNLVEMYL 360
 DB 299 WETGKGLERAVAKYSMEDAKYTFELGKEFFPMEALAKIGOSVWDVSRSTGNLVEMYL 358
 QY 361 RYAVERNELAPNKPDEEYRRRLRTYLLGYVKEPERGMENTIYDFRCLYSIITVHN 420
 DB 359 RYAVERNELAPNKPDEEYRRRLRTYLLGYVKEPERGMENTIYDFRCLYSIITVHN 418
 QY 421 VSPDTEREGCKNVDAPIVGYKFCDFPGFISILGELITMROEIKKKKATIDPIEK 480

Db	419	VSPTLNRKCNKEIDVAPQÜGHRCCKDPGFIISLJONLLEEROKIKRMRKESKDPEKK	478
Qy	481	MLDYRQRAVKLHANSYGYMGYPRAWYSKCAESVTAMGRHYLTMTKEIEEKGKVL	540
Db	479	LIDYRQRAIKRLANSYGYGYAKARWYCKEACSAESTAMGRÖYIDLVRELESS -GPKVL	537
Qy	541	YADIDGEGYATIPGKRPETIKKAKAEPLKTIYNSKIPGLLEIEBSFYLRGFFVAKKRAVI	600
Db	538	YIIDDGUYATIPGAKPMEIKERAKFMEYINSKIPGLLEIEBSFYARGFFVAKKRYALI	597
Qy	601	DESGRIITRGLEYVRROMSEIAKTOAKVLEALIKEDSVKAEVIYDVVEETAKTOYPL	660
Db	598	DEEKIYTRGIEIYRROMSEIAKTOAKVLEALIKESVDVAKYNAVEYIEKISKIETIP	657
Qy	661	EKLVIHEOITKDLSEYKAGPHVAIAKRLAKGIVRPGTIISYIVLRSGSKISDRYILL	720
Db	658	EKLVIYEQITPRLSEYKAGPHVAIVAKRLAKGVKAPGVIYIGIYLRGOGPISKRAIAI	717
Qy	721	SEYDPKHKHXDPDYIENOVYPAVLRLLAEAFGRKKELOKQSSQVGLDAMLK	773
Db	718	EEBPQKHKHXDAEYIENOVYPAVERLRLAEAFGRKKELOKQKQVGLGAMLK	770

```

Db 479 LLDYRQAIKILKLANSFYGYGAKARWYCKECSAESYAWMRKYIELWKKLEEFGRKVL 538
QY 541 YADTDGYATIPGKPKPTIKKAKKFLKNSLPGLELEGEYGRGFEVAKKRAVI 600
Db 539 YIDTDGATITPGGESEIKKALEPKYKNSLPGLELEGEYGRGFEVAKKRAVI 598
QY 601 DEGRITTRGLEVVRDMSIAKETQAKVLEILKEDSVKAEIYKDVVEIAIKYQVPL 660
Db 599 DEGRKVTTRGLEIVRRDMSIAKETQAKVLEILKEDSVKAEIYKDVVEIAIKYQVPL 658
QY 661 EKLVHQRITKDSSEYKATIPHAIAKRLAKIGKAPRGITISYIVLRGSKISDRVIL 720
Db 659 EKALVEQITRPLEHKYKAIPHAIVAAKRLAKIGKAPRGITISYIVLRGSKISDRVIL 718
QY 721 SEYDPKKHKYDPYIENQVLPVAVLRILEAFGRKEDLKQSSKQVGLDAML 772
Db 719 EEDPKHKHKYDAEYIENQVLPVAVLRILEAFGRKEDLKQSSKQVGLDAML 770

```

RESULT 8

POL_PYRSD STANDARD; PRT; 1312 AA.

```

AC 051334; 051335; 051336;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (Deep vent DNA polymerase) [Contains:
GN Endonuclease P1-Papi (EC 3.1.-.-) (Psp-GDB pol Intein)].
OS Pyrococcus sp. (strain GB-D).
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=69013;
RN [1]
RX MEDLINE=94094330; PubMed=8269515;
RA Xu M.-Q., Southworth M.W., Mersha F.B., Hornstra L.J., Peller F.B.;
RT In vitro protein splicing of purified precursor and the
RL Cell 75:1371-1377(1993).
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY.
CC -1- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE
INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY
ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE
INTEIN.
CC -1- CATALYTIC ACTIVITY: N deoxyucleoside triphosphate - N diphosphate
+ [DNA](N).
CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
FOLLOWED BY PEPTIDE LIGATION.
CC -1- BIOCHEMISTRY: Used in the PCR method because of its high
thermostability and low error rate. Sold by New England Biolabs.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
ENDONUCLEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC EMBL: U00707; AAA67130.1; -
DR EMBL: U00707; AAA67131.1; -
DR EMBL: U00707; AAA67132.1; -
DR HSSP: P56689; ITGO.
DR REBASE: 2619; P1-Papi.
DR InterPro: IPR002064; DNA_pol_B.
DR InterPro: IPR003586; Hntc.
DR InterPro: IPR003587; HntcN.

```

```

DR InterPro: IPR002203; Intein.
DR InterPro: IPR004042; Intein_endonuc.
DR Pfam: PF00136; DNA_pol_B_2.
DR Pfam: PF03104; DNA_pol_B_exo_1.
DR PRINTS: PR00106; DNAPOLB.
DR PRINTS: PR00379; INTEIN.
DR SMART: SM00305; Hntc; 1.
DR SMART: SM00306; HntcN; 1.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B_1.
DR PROSITE: PS00881; PROTEIN_SPLICING_1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
KW Protein splicing; Intein homing.
FT CHAIN 1 492 DNA POLYMERASE, 1ST PART.
FT CHAIN 493 1029 ENDONUCLEASE P1-PSPI.
FT CHAIN 1030 1312 DNA POLYMERASE, 2ND PART.
SQ SEQUENCE 1312 AA; 152852 MW; B62518805641D26A CRC64;

```

Query Match

Best Local Similarity 72.9%; Score 2934.5; DB 1; Length 1312;
Matches 597; Conservative 86; Mismatches 87; Indels 539; Gaps 3;

```

QY 1 MIFDTYITKDKGPIIRIFKENGKFKIELDPHOPYIALKDKSDAIDEIKAKGERRG 60
Db 1 MILDADYITEDGPIIRIFKENGKFKIELDPHOPYIALKDKSDAIDEIKAKGERRG 60
QY 61 KIVRVADVAVKKKFGROVYKKLIFEHPOVPALRGKREHNAVYIDYEDYIPAKRY 120
Db 61 KIVRIIDAEKVRKKKFGRLPEVWRKLFEPHODPALRDKRHSNAVYIDYEDYIPAKRY 120
QY 121 LIDKGLIPMGDEDLKLMADFIEFYHEGDEFGKGLIMISYADEEARVITWKNDLPY 180
Db 121 LIDKGLIPMGDEDLKLMADFIEFYHEGDEFGKGLIMISYADEEARVITWKNDLPY 180
QY 181 VDVVSNREMIKRFVQIVRKPDVLTYYNGDNFDLPYLKRAKRGVLLGRDKNEHE 240
Db 181 VDVVSNREMIKRFVQIVRKPDVLTYYNGDNFDLPYLKRAKRGVLLGRDKNEHE 240
QY 181 VEYVSSEREMIKRFLVIREKPDVLTYYNGSFDLPYLKRAKRGVLLGRDKNEHE 238
Db 181 VEYVSSEREMIKRFLVIREKPDVLTYYNGSFDLPYLKRAKRGVLLGRDKNEHE 238
QY 241 PKIHRGDSFAVAVIKIRIHFDLPVVRRTNLTYYLLEAVYEVNLKTSKLSAEETIAI 300
Db 241 PKIHRGDSFAVAVIKIRIHFDLPVVRRTNLTYYLLEAVYEVNLKTSKLSAEETIAI 300
QY 239 PKMGRGDMAVAIKIRIHFDLVHVRITNLTYYLLEAVYEVNLKTSKLSAEETIAI 298
Db 239 PKMGRGDMAVAIKIRIHFDLVHVRITNLTYYLLEAVYEVNLKTSKLSAEETIAI 298
QY 301 WETESNKIAIQYSMEDARATYELGKEFPMEALAKLIGQSVADYSRSSTGNLVEMYL 360
Db 301 WETESNKIAIQYSMEDARATYELGKEFPMEALAKLIGQSVADYSRSSTGNLVEMYL 360
QY 299 WETGKLEIRYAKYSMEDAKYTELGRFPFMEALSRVGOPLMDVSRSSSTGNLVEMYL 358
Db 299 WETGKLEIRYAKYSMEDAKYTELGRFPFMEALSRVGOPLMDVSRSSSTGNLVEMYL 358
QY 361 RYAVERNELAPNKPDEEYRRLRTYVLGKYVERPGRMENTYIDFRLYPSITVTN 420
Db 361 RYAVERNELAPNKPDEEYRRLRTYVLGKYVERPGRMENTYIDFRLYPSITVTN 420
QY 359 RKAVERNELAPNKPDEEYRRLRTYVLGKYVERPGRMENTYIDFRLYPSITVTN 418
Db 359 RKAVERNELAPNKPDEEYRRLRTYVLGKYVERPGRMENTYIDFRLYPSITVTN 418
QY 421 VSPDTEREGCKYVDVAPVIGYFKCDFPGFISIGELITMROEIKKKAKATIDIERK 480
Db 421 VSPDTEREGCKYVDVAPVIGYFKCDFPGFISIGELITMROEIKKKAKATIDIERK 480
QY 419 VSPDTEREGCKYVDVAPVIGYFKCDFPGFISIGELITMROEIKKKAKATIDIERK 478
Db 419 VSPDTEREGCKYVDVAPVIGYFKCDFPGFISIGELITMROEIKKKAKATIDIERK 478
QY 481 MLDYKORAVK----- 490
Db 481 MLDYKORAVK----- 490
QY 479 MLDYRQAIKILKLANSFYGYGAKARWYCKECSAESYAWMRKYIELWKKLEEFGRKVL 538
Db 479 MLDYRQAIKILKLANSFYGYGAKARWYCKECSAESYAWMRKYIELWKKLEEFGRKVL 538
QY 491 ---LHA----- 493
Db 491 ---LHA----- 493
QY 494 ----- 493
Db 494 ----- 493
QY 599 VEATGEDVKTGDLAVPRSVNLPKERNLIVLLNLSPEETEDILITPVKGRKFK 658
Db 599 VEATGEDVKTGDLAVPRSVNLPKERNLIVLLNLSPEETEDILITPVKGRKFK 658
QY 494 ----- 493
Db 494 ----- 493
QY 659 GMLRTLWIFGEKRVATASRYLRHLNGLYIRLRKIGYDIDKEGLEKRYLYEKLVYD 718
Db 659 GMLRTLWIFGEKRVATASRYLRHLNGLYIRLRKIGYDIDKEGLEKRYLYEKLVYD 718
QY 494 ----- 493
Db 494 ----- 493

```

```

Db 719 VRYNGKREYIIEFNAVDVISMPEELKEWRIGTRNGFMGTVEIDDEDAKLLGYV 778
Qy 494 ----- 493
Db 779 SEGSAKKKNOTGWSYVRLYNENDEVLDDMEHLAKKFGCKVKKGNVETIPKMAVYI 838
Qy 494 ----- 493
Db 839 FESLOCTIAENKRVPEVIFTSSKGVRAFLBEGYFIGDDDVHRSKRVRSTSELLVNGIV 898
Qy 494 ----- 493
Db 899 LLNLSGVSAIKLGYDSGVRYVNEELKTEYRKKKNVSHIVKDKLKEFGKVFQK 958
Qy 494 ----- 493
Db 959 NISYKRELYENGLDREKAKRIEMLLNGDIVDRAVEIKREYDGYVLDSDVDENF 1018
Qy 494 ----- 493
Db 1019 LAGFGLVAHNSYGYGYAKARWCKCAESVTAMGREYIEFYAKLEEKGEFVLYID 1078
Qy 544 TDGFYATIPGEKPEPTIKKAKFEKLYINSKLPGLLELEYEGFYLRGFVAKRVAIDEE 603
Db 1079 TDGLYATIPGAKPEEIKKALEFVDYINAKLPGLLELEYEGFYLRGFVAKRVAIDEE 1138
Qy 604 GRITRGLEEVARRDSEIATKETOAKVLEALLKEDSVKAVETVKNVVEIAKYOYPLKEL 663
Db 1139 GKITRGLEIVARRDSEIATKETOAKVLEALLKEDSVKAVETVKNVVEIAKYOYPLKEL 1198
Qy 664 VIHEQITKDLSEYKAIQPHVAIAKRLAKGIVRPGTIISYIVLNGSKISDRVILLSEY 723
Db 1199 VIYEGITRPHLEHYKAIQPHVAIAKRLAKGIVRPGTIISYIVLNGSKISDRVILLSEY 1258
Qy 724 DPKKKYPODYIENOVPAVLRILEAGYRKEDLKYSSKVOYGDAML 772
Db 1259 DLKHKYDAEYIENOVPAVLRILEAGYRKEDLKYSSKVOYGDAML 1307

RESULT 9
DPOL_PRRHO STANDARD: PRT: 1235 AA.
AC 059610:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) [contains: Pho pol intein (Pho pol I
inleln)].
GN POL OR PH1947 OR PHBTO47.
OS Pyrococcus horikoshii.
OC Archaea: Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
Yanamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamaya M., Ohnuku Y.,
Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
Masuchi Y., Shizuya H., Kikuchi H.;
RA "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RT DNA Res. 5:55-76(1998).
RL
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
+ [DNA](N).
CC -1- PM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000007; BAA31074.1; -
DR HSSP: P56689; INGO.
DR InterPro: IPR002064; DNA_POL_B.
DR InterPro: IPR003586; HincC.
DR InterPro: IPR003587; HincN.
DR InterPro: IPR002203; Intein.
DR InterPro: IPR004042; Intein_endonuc.
DR Pfam: PF00136; DNA_POL_B; 2.
DR Pfam: PF01104; DNA_POL_B_exo; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00305; HincC; 1.
DR SMART: SM00306; HincN; 1.
DR SMART: SM00486; POLB; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
DR PROSITE: PS00881; PROTEIN_SPLICING; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Autocatalytic cleavage; Protein splicing;
KW Complete proteome.
FT CHAIN 1 492 DNA POLYMERASE, 1ST PART (POTENTIAL).
FT CHAIN 493 952 PHO POL INTEIN (POTENTIAL).
FT CHAIN 953 1235 DNA POLYMERASE, 2ND PART (POTENTIAL).
SQ SEQUENCE 1235 AA; 143086 MW; 73CC7AA14873CCE4 CRC64;

Query Match 71.7%; Score 2886; DB 1; Length 1235;
Best Local Similarity 46.5%; Pred. No. 6,5e-15;
Matches 574; Conservative 107; Mismatches 89; Indels 464; Gaps 5;

Qy 1 MIEDDYITTKDKPIIRIFKENGSEFKTELDPHQPIYVALLKDDSAIDEIKAKGERHG 60
Db 1 MILDADYITTEDCKPIIRIFKENGSEFKTELDPHQPIYVALLKDDSAIDEIKAKGERHG 60
Qy 61 KIYRVYDAVKKKKKLGSDVEYWKLFEPHPOVPAALRKIRHPAVIDIEYDIFARY 120
Db 61 KVRIVETEKIQKRLGPRIEYWKLYLEHPQVPAIRKIRHPAVVDFIEYDIFARY 120
Qy 121 LIDKGLIMEGDEELKLAFFDIETFYHGEDEFGKEIIMISVADDEEAVITWKNDIPY 180
Db 121 LIDKGLITMEGNEKLTFLAVDIEITLYHGEDEFGKPPVIMISVADDEEAKVITWKNDIPY 180
Qy 181 VDVSNEREMIKRFVQIVREKDPDVLITYNGDNFDPYLIRKRAEKLGVTLGRDKHEPE 240
Db 181 VEVVSSEREMIKRLIRVIRKEKDPDVIITYNGDNFDPYLIRKRAEKLGVTLGRDKHEPE 240
Qy 241 PKIRHMGSPAVEIKRGRIHFDLPFVYVRTINLPYTLLEAVYAVLGKRSKGAETIAI 300
Db 241 PKMOKMGDSLAVEIKRGRIHFDLPFVYVRTINLPYTLLEAVYAVLGKRSKGAETIAI 300
Qy 301 WETESMKKLQVSMEDARATYELGKEFFPMAELAKLIGSVMPVSSSTGNLEWYLL 360
Db 301 WETGSLGLERVAKYSMEDAKVITELGREFFPMAQALRVGPVMPVSSSTGNLEWYLL 360
Qy 361 RYAVERNELAPNKPDDEEYRRRLRTTYLGYYKVEBERGLMENTYLDRECLYPSITVTHN 420
Db 361 RYAVERNELAPNKPDDEEYRRRLRTTYLGYYKVEBERGLMENTYLDRECLYPSITVTHN 420
Qy 421 VSPDTLREGECKNRYAVIYGVKFKKDPGPFPSILGELITRQKIKKMKATIDPIEKK 480
Db 421 VSPDTLREGECKNRYAVIYGVKFKKDPGPFPSILGELITRQKIKKMKESKDPVEKK 480
Qy 492 ----- 491
Db 492 ----- 491

```

```

Db 539 VKDLKALSFNRETAKSELKKVYKALIRHRSYSGKYVSIKLKSGRIKITSGHSLFSVKNKGL 598
QY 492 ----- 491
Db 599 VKVAGDELKPGDLVYVPGRLKLPESKQVNLVVELLKLPEEETSNIVAMIPVKGRKNFEK 658
QY 492 ----- 491
Db 659 GMLKTLVIFGEGERPRTAGRYLKLHLERLGYVYKLRRCGEVLDMESLKYRYKLEYTLKN 718
QY 492 ----- 491
Db 719 LKYNAGSRATWVEFNSLRDYYSLMPLELKEWITGEPRGKIGTFIDVDSFAKLGYI 778
QY 492 ----- 493
Db 779 SSGDVEKDRVYKFSKDDQNVLEDIAKLAKLFGKYRGRGYIEVSGKISHAIFRYLAKGR 838
QY 494 ----- 493
Db 839 IPEFIPTSPMDIKVAFKLGNGNAEELTFSTKSELLVNLILLINSIGVSDIKIEHEKV 898
QY 494 ----- NSYGYM 500
Db 899 YRYVYINKKSSNGDIVDSVESIEVEKYEGYVDLSVEDNENFLVGFLLYAHNSYGY 958
QY 501 GYPRKAWYSKCAESVTAMGRHYTEMITKEIEKFGKVLADTGFYATIPGKRP-ETI 559
Db 959 GYKARWYKCEKCAESVTAMGRHYTEMITKEIEKFGKVLADTGFYATIPGKRP-ETI 559
QY 560 KKKAKELKATNGKLPGLLEVEGFLRGFPYAKKRYAVIDEGRITTTGLEGVYRDM 619
Db 1018 KRRLEFVDYINSKLPGLLEVEGFLRGFPYAKKRYAVIDEGRITTTGLEGVYRDM 1077
QY 620 EIAKETQAKYLEALIKEDSYEKAVEIYKQVVEIAKQVLEKVIHEOTKDLSEKAI 679
Db 1078 EIAKETQAKYLEALIKEDSYEKAVEIYKQVVEIAKQVLEKVIHEOTKDLSEKAI 679
QY 680 GPHVAIAKRLANGIKVYRPTIISYIVLRGSKISDRVILLSEYDPKHKYDPDYIENQ 739
Db 1138 GPHVAIAKRLANGIKVYRPTIISYIVLRGSKISDRVILLSEYDPKHKYDPDYIENQ 739
QY 740 VLPVAVLILAEAFGRKEDLKYOSKQVGLDAMIK 773
Db 1198 VLPVAVLILAEAFGRKEDLKYOSKQVGLDAMIK 773
RESULT 10
POLYMERASE STANDARD: PRT; 1523 AA.
AC 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) (Pol Tfu) [Contains: Endonuclease PI-TfuI
DE (EC 3.1.-.-) (Tfu pol-1 Intein); Endonuclease PI-TfuII (EC 3.1.-.-)
DE (Tfu pol-2 Intein)].
GN POL.
OS Thermococcus fumicolans.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI_TaxID:46540;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ST557;
RA Camdon M., Querellou J.;
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP CHARACTERIZATION OF INTEINS.
RC STRAIN-ST557;
RA MEDLINE-20112788; PubMed-10644683;
RA Saves I., Ozanne V., Dietrich J., Masson J.-M.;
RT "Inteins of Thermococcus fumicolans DNA polymerase are endonucleases

```

```

RT with distinct enzymatic behaviors."
RL J. Biol. Chem. 275:2335-2341(2000).
CC -1- FUNCTION: PI-TfuI recognizes and cleaves a minimal sequence of 16
CC base pairs (bp) on supercoiled DNA with either Mn(2+) or Mg(2+) as
CC cofactor. It cleaves linear DNA only with Mn(2+) and requires a
CC 19-bp minimal recognition sequence. The optimal temperature for
CC activity is 70 degrees Celsius.
CC -1- FUNCTION: PI-TfuII is a highly active homing endonuclease using
CC Mg(2+) as cofactor. Its minimal recognition and cleavage site is
CC 21 bp long either on linear or circular DNA substrates. Its
CC endonuclease activity is strongly inhibited by the 3' digestion
CC product, which remains bound to the enzyme after the cleavage
CC reaction. The optimal temperature for activity is 70 degrees
CC Celsius.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + {DNA}(N).
CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE TWO INTERVENING REGION
CC (INTEINS) FOLLOWED BY PEPTIDE LIGATION.
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION, BELONGS TO THE HOMING
CC ENDONUCLEASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC
CC EMBL: Z69882; CAA93738.1; -
CC HSSP: P56689; ITGO.
CC DR REBASE: 4500; PI-TfuI.
CC DR REBASE: 4501; PI-TfuII.
CC DR InterPro: IPR002064; DNA_POL_B.
CC DR InterPro: IPR003586; HincC.
CC DR InterPro: IPR003587; HincN.
CC DR InterPro: IPR002203; Intein.
CC DR InterPro: IPR004042; Intein_endonuc.
CC DR Pfam: PF00136; DNA_pol_B_3.
CC DR Pfam: PF03104; DNA_pol_B_exo_1.
CC DR PRINTS: PR00379; INTEIN.
CC DR SMART: SM00305; HincC_2.
CC DR SMART: SM00306; HincN_2.
CC DR SMART: SM00486; POLBC_1.
CC DR PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
CC DR PROSITE: PS00881; PROTEIN_SPLICING_2.
CC KW Transferase; DNA-directed DNA polymerase; DNA replication;
CC DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
CC Protein splicing; Intein homing; Magnesium; Manganese.
CC FT CHAIN 1 406
CC FT CHAIN 407 766
CC FT CHAIN 767 900
CC FT CHAIN 901 1282
CC FT CHAIN 1283 1523
CC FT CHAIN 1523 175917 MW: 7A2AC8236BF2E5F5 CRC64;
CC SEQUENCE
QY Query Match 69.8%; Score 2809; DB 1; Length 1523;
QY Best Local Similarity 39.0%; Pred. No. 1,8e-150;
QY Matches 593; Conservative 86; Mismatches 91; Indels 752; Gaps 4;
Db 1 MIFDPTITKDGKPIRIKKEGKPEIIEPFGQYIYALKDSDAIDEIKAIKGRHG 60
Db 1 MILDPTITEDGKPIYRVKKEGKPEIIEPFGQYIYALKDSDAIDEIKAIKGRHG 60
QY 61 KIVRVVDAVKKKKLGRVYWKLFIEHPQDVPAALRGKIREHPAVIDIYEDIPAKRY 120
Db 61 TIVRVVDAVKKKKLGRVYWKLFIEHPQDVPAALRGKIREHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLIPGDEELKMAFDIETFYHEDGFGKGIIMISVADDEAVITWKNIDLPY 180

```

Db 121 LIDKLLIMEGDELMKLAFLDIETLYHEGEFEAPGPIIMISYADEGARVITWKKIDLPY 180
 Qy 181 VDVASNEREMIKRFVOIAREKDPDLITYNGDNFPLPYLIKRAEKLGTLLIGRDEKHE 240
 Db 181 VDVASNEREMIKRFVKVKEKDPDLITYNGDNFPAFLKRSKSEKLGKFTLIGRGS--E 238
 Qy 241 PKIRHMGDSFAVEIKGRHFDLPVVRRTINLPYTLAEVAEYAGTKRSKLGAEIAT 300
 Db 229 PKIQMGDRFAVEYKGRHFDLPYIRHTINLPYTLAEVAEYAGTKRSKLGAEIAT 298
 Qy 301 WETESMKKLQYSMEDARATTELCKEFPFMAELAKLIGOSWDSVSSSTGNLVEWYL 360
 Db 299 WETGGLERVAARYSMEDAKVYTELKEFPFMAELAKLIGOSWDSVSSSTGNLVEWYL 358
 Qy 361 RFAVERNELAPKPDPEEEXRRRLRTTYLGYYKEPERGIMETITLDRPC----- 410
 Db 359 RFAVERNELAPKPDPEEEXRRRLRTTYLGYYKEPERGIMETITLDRPC----- 410
 Qy 411 ----- 410
 Db 418 GKGVNISEVREGDYVLGIDGQKVQVWEXDYEGELVNINGLKCTPNHKLPRVARTERQ 477
 Qy 411 ----- 410
 Db 478 TAIRDLSAKSFLTKKVKGLITTPLEKIGIEREDVEEELKGLAGILAEGLTLRK 537
 Qy 411 ----- 410
 Db 538 DVEYDSRGGKRVSHQYRVELITVGAQEDDQORLIVYIFERLFGVTPSYRKKNATLTF 597
 Qy 411 ----- 410
 Db 598 KVAKKEYLVRREIMDGIENLHAPSVLGPFEGDSVANKRVTVVNOGTNEMKIEVYS 657
 Qy 411 ----- 410
 Db 658 KLNKLGIPHRRYTYDYTEREKTMTTHILEIAGRDGLLFTQIVGISTEKMALEAIR 717
 Qy 411 ----- 410
 Db 718 NREVRLENNAFYTLADTAKTEYKGVYDLTEGTPPYFANGILTHNSLPSITLISHN 777
 Qy 421 VSPDTLEBEGCKNDYDAPVIGYKCKDPGFPISILGELITMRQEKKKMATIDPIEK 480
 Db 778 VSPDTLEBEGCKNDYDAPVIGYKCKDPGFPISILGELITMRQEKKKMATIDPIEK 837
 Qy 481 MLDYQORAVKLANHYGYMGYKPKAWYKSKCEASVTYAMGRHYIMTKEIEKGFVYL 540
 Db 838 LLDYQORAVKLANHYGYMGYKPKAWYKSKCEASVTYAMGRHYIMTKEIEKGFVYL 897
 Qy 541 YAD----- 543
 Db 898 YADVTGTGTEVTIRNGRIEYPIKLEPERVDHRYGEKEVYCVLGEVYALTLDNKRLYWK 957
 Qy 544 ----- 543
 Db 958 KVPYVRRKTDKRIYRWFTNSWYLDYEDHSLIGYMTSKYKPGKPLKERLVEYKPEEL 1017
 Qy 544 ----- 543
 Db 1018 GGVKSLITPNPARIKANDIAVKLWELIGLVGDSNGWGSWMARYVYGLSCGLDKA 1077
 Qy 544 ----- 543
 Db 1078 EIERKVLPLREASVYSNYDKSKKGDVSIISKWLAGPMVKYFKDENGKALPSEFMDLP 1137
 Qy 544 ----- 543
 Db 1138 REYIEAFILGELFSDAGTVSLRGIPEIRLTSYVNRSLDAVKRLMLVGVNSLFTETKPN 1197
 Qy 544 ----- 543
 Db 1198 RYLEKESGTHSHIRIKNNKRRPADRIGFLIDRKSTKLSENLGHTNNKRAYKYDEDLVYP 1257

Qy 544 -----TDGFATIPGKPEPTIKKRAEFLKYN 571
 Db 1258 KRIEETTDGYYVYDIEVEGTHFFPANGILVHTDGFATIPADAEYKAKKREPLANTIN 1317
 Qy 572 SKPLGLELEVEGYFLRGFEVAKKRRVAVDEBGRITTRGLEVVRDMSEIAKETQAVYLE 631
 Db 1318 PLPLGLELEVEGYFLRGFEVAKKRRVAVDEBGRITTRGLEVVRDMSEIAKETQAVYLE 1377
 Qy 632 ALKEDSVKAEVYKDVVEELAKYQVPLEKVIYHEQITKDLSEYKALGPHVAIKRLAA 691
 Db 1378 ALRKGVDVEEAARYKVEYTELKSKYVPEKLVYHEQITRELKDYKAGSPHVAIKRLAA 1437
 Qy 692 KGIVRPGTIIISYIYLRGSGKISDRIYLLSEYDPKKHKYDPDYIENQVLPVLRILEAF 751
 Db 1438 RGIKVRPGTIIISYIYLRGSGKISDRIYLLSEYDPKKHKYDPDYIENQVLPVLRILEAF 1497
 Qy 752 GYKEDLKYOSSKQVGLDAMLK 773
 Db 1498 GYKEDLKYOSSKQVGLDAMLK 1519
 RESULT 11
 DPOL_PYRO STANDARD: PRT: 1671 AA.
 AC P77933;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA polymerase (EC 2.7.7.7) [contains: Endonuclease PI-PKOI
 DE (EC 3.1.-.-) (Pko pol-1 intein) (IVS-A); Endonuclease PI-PKOII
 DE (EC 3.1.-.-) (Pko pol-2 intein) (IVS-B)].
 GN POL.
 OS Pyrococcus kodakaraensis.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
 OX NCBI_TaxID=69014;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC STRAIN-KOD1;
 RX MEDLINE=98027387; PubMed=9361436;
 RA Takagi M., Nishiohara M., Kakinaka H., Kitabayashi M., Inoue H.,
 RA Kawakami B., Oka M., Imanaka T.;
 RT "Characterization of DNA polymerase from Pyrococcus sp. strain KOD1
 RL Appl. Environ. Microbiol. 63:4504-4510(1997).
 RN [2]
 RP CHARACTERIZATION OF INTEINS.
 RC STRAIN-KOD1;
 RX MEDLINE=98416198; PubMed=9742242;
 RA Nishiohara M., Fujiwara S., Takagi M., Imanaka T.;
 RT "Characterization of two intein homing endonucleases encoded in the
 RL DNA polymerase gene of Pyrococcus kodakaraensis strain KOD1.";
 CC Nucleic Acids Res. 26:4409-4412(1998).
 CC -1- FUNCTION: INTEIN ENCODED ENDONUCLEASES ARE THOUGHT TO MEDIATE
 CC INTEIN MOBILITY BY SITE-SPECIFIC RECOMBINATION INITIATED BY
 CC ENDONUCLEASE CLEAVAGE AT THE "HOMING SITE" IN GENE THAT LACK THE
 CC INTEIN. PI-PKOI RECOGNIZES 5'-GATTGATGCCGTACC-3' AND PI-PKOII
 CC RECOGNIZES 5'-CAGCTACTACGGTTAC-3'. BOTH ARE THERMOSTABLE.
 CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
 CC + [DNA](N).
 CC -1- PRT: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 CC FOLLOWED BY PEPTIDE LIGATION (PORENTIAL).
 CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
 CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
 CC ENDONUCLEASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@slb-slb.ch).

DR EMBL: D29671; BAA06142.2; .
DR HSSP: P56689; ITGO.
DR REBASE: 3792; PI-PKOI.
DR REBASE: 3793; PI-PKOI.
DR InterPro: IPR002064; DNA_pol_B.
DR InterPro: IPR003586; HincC.
DR InterPro: IPR003587; HincN.
DR InterPro: IPR002203; Intein.
DR InterPro: IPR004042; Intein_endonuc.
DR Pfam: PF00136; DNA_pol_B_3.
DR Pfam: PF03104; DNA_pol_B_exo_1.
DR PRINTS: PR00379; INTEIN.
DR SMART: SM00305; HincC_2.
DR SMART: SM00306; HincN_2.
DR SMART: SM00486; POLBc_1.
DR PROSITE: PS00116; DNA_POLYMERASE_B_1.
DR PROSITE: PS00881; PROTEIN_SPLICING_2.
KW Transferrase; DNA-directed DNA polymerase; DNA replication;
M DNA-binding; Hydrolase; Endonuclease; Autocatalytic cleavage;
N Protein splicing; Intron homing.
FT CHAIN 1 406 DNA POLYMERASE, 1ST PART.
FT CHAIN 407 766 ENDONUCLEASE PI-PKO I.
FT CHAIN 767 851 DNA POLYMERASE, 2ND PART.
FT CHAIN 852 1388 ENDONUCLEASE PI-PKO II.
FT CHAIN 1389 1671 DNA POLYMERASE, 3RD PART.
SQ SEQUENCE 1671 AA; 193490 MW; 4A17F1C8120EE455 CRC64;

Query Match 69.1%; Score 2782; DB 1; Length 1671;
Best Local Similarity 36.0%; Pred. No. 6.9e-149;
Matches 602; Conservative 82; Mismatches 86; Indels 900; Gaps 4;

QY 1 MIDDTYITKDGKPIIRIFKENGFEKIELDPHOPIYALLKDDSAIDEIKATKGRHG 60
1 MIDDTYITKDGKPIIRIFKENGFEKIELDPHOPIYALLKDDSAIDEIKATKGRHG 60
1 MIDDTYITKDGKPIIRIFKENGFEKIELDPHOPIYALLKDDSAIDEIKATKGRHG 60
QY 61 KIYRVDAVAKKKKFLGRDVEYWKLFIEHPQDVALRGKIREHPAVIDIYEDIPAKRY 120
61 KIYRVDAVAKKKKFLGRDVEYWKLFIEHPQDVALRGKIREHPAVIDIYEDIPAKRY 120
61 KIYRVDAVAKKKKFLGRDVEYWKLFIEHPQDVALRGKIREHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLIPMGDEBELKLMADIEFYHEGDEPGKELIMISYDEEARITTKNDLPY 180
121 LIDKGLIPMGDEBELKLMADIEFYHEGDEPGKELIMISYDEEARITTKNDLPY 180
121 LIDKGLIPMGDEBELKLMADIEFYHEGDEPGKELIMISYDEEARITTKNDLPY 180
QY 181 VDVVSNEREMIKRFVOIVREKDPVLITYNGDNFDLPYLTKRAEKLGVTLTGRDKEHP 240
181 VDVVSNEREMIKRFVOIVREKDPVLITYNGDNFDLPYLTKRAEKLGVTLTGRDKEHP 240
181 VDVVSNEREMIKRFVOIVREKDPVLITYNGDNFDLPYLTKRAEKLGVTLTGRDKEHP 240
QY 241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 300
241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 300
241 PKIHRMGDSFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 300
QY 239 PKIQRMGDRFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 298
239 PKIQRMGDRFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 298
239 PKIQRMGDRFAVEIKGRIFHDLFPVVRRTINLPYTLLEAVYEAVLGKTSKLGAEETAI 298
QY 301 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 360
301 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 360
301 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 360
QY 299 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 358
299 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 358
299 WEFEESMKKIAQSMEDARTYELGKEFFPMEALAKLIGOSWVDVSRSTGLVENVYL 358
QY 361 RVAVENNELAPNKPDEEYRRLRTTYLGKYEKPERGLMENTIYDFRC----- 410
361 RVAVENNELAPNKPDEEYRRLRTTYLGKYEKPERGLMENTIYDFRC----- 410
361 RVAVENNELAPNKPDEEYRRLRTTYLGKYEKPERGLMENTIYDFRC----- 410
QY 359 RKAYENNELAPNKPDEELARR--RQSYEGGYAKPERGLMENTIYDFRC----- 417
359 RKAYENNELAPNKPDEELARR--RQSYEGGYAKPERGLMENTIYDFRC----- 417
359 RKAYENNELAPNKPDEELARR--RQSYEGGYAKPERGLMENTIYDFRC----- 417
QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 418 GKGIINISEQEGDYVLIGDMQVRKRWEDYKGEIVNINGLKTPNHKLPVATKNERG 477
418 GKGIINISEQEGDYVLIGDMQVRKRWEDYKGEIVNINGLKTPNHKLPVATKNERG 477
418 GKGIINISEQEGDYVLIGDMQVRKRWEDYKGEIVNINGLKTPNHKLPVATKNERG 477
QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 478 TRIROSLANSFLTKKVKGLITTPLEIYIGRATSENIPEEEVLKGLAGIILAEGTLRK 537
478 TRIROSLANSFLTKKVKGLITTPLEIYIGRATSENIPEEEVLKGLAGIILAEGTLRK 537
478 TRIROSLANSFLTKKVKGLITTPLEIYIGRATSENIPEEEVLKGLAGIILAEGTLRK 537
QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 538 DVEYFDSSKKRRKRIHQYRVEITIGKDEEFRDRTIYIFERLFGITPISSEKGTNAVTL 597
538 DVEYFDSSKKRRKRIHQYRVEITIGKDEEFRDRTIYIFERLFGITPISSEKGTNAVTL 597
538 DVEYFDSSKKRRKRIHQYRVEITIGKDEEFRDRTIYIFERLFGITPISSEKGTNAVTL 597

QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 598 KVAKKNVYLKKEIMONIESLHPSVLRGFEEDGSVNRVRSIVATQGTKNEMKIKLYS 657
598 KVAKKNVYLKKEIMONIESLHPSVLRGFEEDGSVNRVRSIVATQGTKNEMKIKLYS 657
598 KVAKKNVYLKKEIMONIESLHPSVLRGFEEDGSVNRVRSIVATQGTKNEMKIKLYS 657
QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 658 KLSQLGIPHOYTYOYOENGKDRSRYILEITGKDLILFQTLIGFISERNALLNKALIS 717
658 KLSQLGIPHOYTYOYOENGKDRSRYILEITGKDLILFQTLIGFISERNALLNKALIS 717
658 KLSQLGIPHOYTYOYOENGKDRSRYILEITGKDLILFQTLIGFISERNALLNKALIS 717
QY 411 ----- 410
411 ----- 410
411 ----- 410
QY 718 QRENNLNNNGFYRLSEFNVSTYEYCKVYDLLEGTPIYFANGILTHNSLYPSIITHN 777
718 QRENNLNNNGFYRLSEFNVSTYEYCKVYDLLEGTPIYFANGILTHNSLYPSIITHN 777
718 QRENNLNNNGFYRLSEFNVSTYEYCKVYDLLEGTPIYFANGILTHNSLYPSIITHN 777
QY 421 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 480
421 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 480
421 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 480
QY 778 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 837
778 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 837
778 VSPDTLEREGCKNDYAPIVYKCKDPGFIPIISIGELITMROEIKKKKATIDPLEKK 837
QY 481 MLDYRQRAVLAH----- 493
481 MLDYRQRAVLAH----- 493
481 MLDYRQRAVLAH----- 493
QY 838 LLDYRQRAVLAH----- 897
838 LLDYRQRAVLAH----- 897
838 LLDYRQRAVLAH----- 897
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 898 VSGLEVPSENRKTKALKRYKALIRHDSGKYTTIRLSGRKIKITSGHSLFVRNGEL 957
898 VSGLEVPSENRKTKALKRYKALIRHDSGKYTTIRLSGRKIKITSGHSLFVRNGEL 957
898 VSGLEVPSENRKTKALKRYKALIRHDSGKYTTIRLSGRKIKITSGHSLFVRNGEL 957
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 958 VEYTGDELKPGDLVAVPRRLPELRNHNVLNVELLGTPEETLDIYMTIPVKKKNFK 1017
958 VEYTGDELKPGDLVAVPRRLPELRNHNVLNVELLGTPEETLDIYMTIPVKKKNFK 1017
958 VEYTGDELKPGDLVAVPRRLPELRNHNVLNVELLGTPEETLDIYMTIPVKKKNFK 1017
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1018 GMLRLRMIFGEERKRPRTARLRLHLEDLGYRLKRGYEVLDWLSLKNRYLALVEN 1077
1018 GMLRLRMIFGEERKRPRTARLRLHLEDLGYRLKRGYEVLDWLSLKNRYLALVEN 1077
1018 GMLRLRMIFGEERKRPRTARLRLHLEDLGYRLKRGYEVLDWLSLKNRYLALVEN 1077
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1078 VRYNKKREVLVFNSTRDAVGLMPLKELKEMKIGTLNGRMKRLLEVDESLAKLLGYV 1137
1078 VRYNKKREVLVFNSTRDAVGLMPLKELKEMKIGTLNGRMKRLLEVDESLAKLLGYV 1137
1078 VRYNKKREVLVFNSTRDAVGLMPLKELKEMKIGTLNGRMKRLLEVDESLAKLLGYV 1137
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1138 SEGARKORNPKNKGSYSVLYNEDPEVLDMEBLASRFEGKVRGRNRYEIPKIGYLL 1197
1138 SEGARKORNPKNKGSYSVLYNEDPEVLDMEBLASRFEGKVRGRNRYEIPKIGYLL 1197
1138 SEGARKORNPKNKGSYSVLYNEDPEVLDMEBLASRFEGKVRGRNRYEIPKIGYLL 1197
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1198 FENMGVLAENKRIPEFVFTSPKGVRLAEGYFIDGDGVHPKRLRLSTKSELNOLY 1257
1198 FENMGVLAENKRIPEFVFTSPKGVRLAEGYFIDGDGVHPKRLRLSTKSELNOLY 1257
1198 FENMGVLAENKRIPEFVFTSPKGVRLAEGYFIDGDGVHPKRLRLSTKSELNOLY 1257
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1258 LLNSGVSAVKLGHDSGYRYVINEBLPVKDKKKNAVYSHVIRKVELSEVFGVFOK 1317
1258 LLNSGVSAVKLGHDSGYRYVINEBLPVKDKKKNAVYSHVIRKVELSEVFGVFOK 1317
1258 LLNSGVSAVKLGHDSGYRYVINEBLPVKDKKKNAVYSHVIRKVELSEVFGVFOK 1317
QY 494 ----- 493
494 ----- 493
494 ----- 493
QY 1318 NVPQTFRKWVEDGRDLPKRAQRLSMLIEGDVLDLVESVDVEDYGYVYDLSVEDNENF 1377
1318 NVPQTFRKWVEDGRDLPKRAQRLSMLIEGDVLDLVESVDVEDYGYVYDLSVEDNENF 1377
1318 NVPQTFRKWVEDGRDLPKRAQRLSMLIEGDVLDLVESVDVEDYGYVYDLSVEDNENF 1377
QY 494 ----- 543
494 ----- 543
494 ----- 543
QY 1378 LVGFGVLAHNSYGYGVARARWYCKEASSTYANGREITWTIEIEKKGFKVYSD 1437
1378 LVGFGVLAHNSYGYGVARARWYCKEASSTYANGREITWTIEIEKKGFKVYSD 1437
1378 LVGFGVLAHNSYGYGVARARWYCKEASSTYANGREITWTIEIEKKGFKVYSD 1437
QY 544 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 603
544 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 603
544 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 603
QY 1438 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 1497
1438 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 1497
1438 TDGFYATIPGKPEITIKKRAKEFLKYNKLPDLLEVEGEFVLRGFFVAKKRYAVIDE 1497
QY 604 GRITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 663
604 GRITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 663
604 GRITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 663
QY 1498 GKITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 1557
1498 GKITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 1557
1498 GKITRGLLEVVRDMSLAKETQAKVLEALIKEDSVKAVEIYKDVVEIYAKYQVPLEK 1557
QY 664 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 723
664 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 723
664 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 723
QY 1558 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 1617
1558 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 1617
1558 VTHEQITKDLSEKAGPHVATKRLAAGKIRKPTIISYIVLRSSKISDPVILLSEY 1617
QY 724 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 773
724 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 773
724 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 773
QY 1618 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 1667
1618 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 1667
1618 DPKKHKYDDPYIENOVLPVAVLILFAFGYRKEDLTQSSKQYGDAMLK 1667

RESULT	12
ID	DPOL_THEG8
STANDARD:	PRF; 1699 AA.
AC	Q9HH84;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	EC 3.1.1.- (Tsp-GE8 pol-1 intein)
DE	(EC 3.1.1.-) (Tsp-GE8 pol-2 intein)
GN	POL OR POL-1.
OS	Thermococcus sp. (strain GE8).
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX	NCBI_TaxId=105583;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Querellou J.J.E., Cambon M.A., Lassongeur F., Barbier G.;
RT	"Thermococcales taxonomy and phylogeny based on the comparative use of
RT	16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
RT	genes.";
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
CC	-1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE
CC	EXHIBITS 3', TO 5' EXONUCLEASE ACTIVITY (BY SIMILARITY).
CC	-1- FUNCTION: PI-TSPGE8I AND PI-TSPGE8II ARE ENDONUCLEASES
CC	(POTENTIAL).
CC	-1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC	+ [DNA] (N).
CC	-1- PFM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC	A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC	FOLLOWED BY PEPTIDE LIGATION.
CC	-1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC	-1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOWING
CC	ENDONUCLEASE FAMILY.
CC	-----
CC	THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC	BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC	THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON
CC	USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC	MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC	ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/
CC	OR SEND AN EMAIL TO license@isb-sib.ch).
CC	-----
DR	EMBL: AJ250333; CAC12850.1; -
DR	InterPro: IPR002064; DNA_POL_B.
DR	InterPro: IPR003586; HincC.
DR	InterPro: IPR003587; HincN.
DR	InterPro: IPR002203; Intein.
DR	InterPro: IPR004042; Intein_endonuc.
DR	Pfam: PF00136; DNA_POL_B; 3.
DR	Pfam: PF03104; DNA_POL_B_exo; 1.
DR	PRINTS: PR00106; DNAPOLB.
DR	PRINTS: PR00379; INTEIN.
DR	SMART: SM00305; HincC; 2.
DR	SMART: SM00306; HincN; 2.
DR	SMART: SM00486; POLB; 1.
DR	PROSITE: PS00881; PROTEIN_SPLICING; 2.
DR	PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
KW	Transferase; DNA-directed DNA polymerase; DNA replication;
KW	DNA-binding; Hydrolase; Nuclease; Exonuclease; Multifunctional enzyme;
KW	Protein splicing; Autocatalytic cleavage; Endonuclease; Intron homing.
FT	CHAIN 1
FT	CHAIN 492 1026
FT	CHAIN 1027 1075
FT	CHAIN 1076 1464
FT	CHAIN 1465 1699
FT	SEQUENCE 1699 AA; 197323 MW; F389BA451F0B12D3 CRC64;

Query Match	68.8%;	Score 2769.5;	DB 1;	Length 1699;
Best Local Similarity	35.5%;	Pred. No. 3.5e-148;		
Matches 602; Conservative	85;	Mismatches 83;	Indels 927;	Gaps 4;

QY	1	MIFPDDYTRKGRPIIEIKKNEKNEFIEIDPHPOPIYALLDDSDAIDEIKAGEERG	60
Db	1	MILDPTDYTEGKQVINYFKKNECFEIEIDRNFEPFYALLDDSDAIEVKKITAKRRG	60
QY	61	KIVRVDAVYKVKKFLIGRDVEYWKLFIEPHODYPALNGKIREHPAVIDIEXYDIPAKRY	120
Db	61	TVVKKRAEKYKKFKFLGRPLEYWKLYFTPHQDPVATIRDKIREHPAVIDIEXYDIPAKRY	120
QY	121	LIDGILPMGEDDELKIMARDIEFFYHGESDFPKGELIMSTADEEAVITWKNIDLPY	180
Db	121	LIDGLLPMGEDDELKIMARDIEFLYHGESDFPKGELIMSTADEGAVITWKNIDLPY	180
QY	181	VDVYVSNEREMKRRVQVIREKDDPVLTITTYNGDNFDPYLTIRAKKIGVTLILGRDKEHPE	240
Db	181	VDVYSTEKEMIKRRLRVYKKEKDDPVLTITTYNGDNFDPYALIKRSEKIGVYFLIGRDS--E	238
QY	241	PKIRHMGDSFAVEIKGAIHEDLPVYVARTINLPYTLLEAVYEAVALGKTYKSUGAEIAAI	300
Db	239	PKIORMGGRFAVEYKGRIHEDLPVIRRTINLPYTLLEAVYEAIFPKPREKYAEIATA	298
QY	301	WETESKKKIAQYSEMDARATYELGKFEFPEMEALAKIGOSVWDSRSSTGNIVBYLL	360
Db	299	WETGEGLEVARYSMEDAKATFELGKFEFPEMQLSGLQSAMDVSNSSTGVLVEMFLL	358
QY	361	RVAYERNELAENKDEEYERRLTATYITLYGKYVKEPERGMENTITTYDIFCLYPSIITVHN	420
Db	359	RKAYERNELAENKDEERELARR--RQSYAGSYKVEPERGMLNNTIVYLDPSLYPSIITVHN	417
QY	421	VSPDTLEBEGCKANDVAPIYGYKPKCKPFGFISILIGELITMAQELTKKKMKATIDPIEK	480
Db	418	VSPDTLLEBEGCKEYDVAQVGHKFCKPFGFIPSLGLDLLEEROKIKRKRATIDPVEKK	477
QY	481	MLDYRORAVYKLAH-----	493
Db	478	LLDYRORAIKILANSILPDEMLPLVYNGRLKLVRIODPVNTMKKQCPLENGTETVLEVS	537
QY	494	-----	493
Db	538	GIEAISFNKRTKIAEIKPVKALIRHRYRGKVYDIKLSGSRNIKVTGHSLEFARDEGELVE	597
QY	494	-----	493
Db	598	VTGGEIKRGDPIAIVRRVNLPERHERINLIEILLGLPPESTDVILTIYVKGRKNFKMG	657
QY	494	-----	493
Db	658	LRTLRWTFEEBQREPTARRYLEHLQKGYVKLAKRAVEIVNKEALRNYRKLEYLAERYK	717
QY	494	-----	493
Db	718	YNGKAREYLVHFNDLRNEIKFMPDELEBKKVGTLNGFRMEPEIYEGEDPAKLLGYVSE	777
QY	494	-----	493
Db	778	GYAKRORNOCKNGMSYSVKIYNNDQRYVLDMEKILASKFPGVRGKNYVEISRKMAVYLE	837
QY	494	-----	493
Db	838	SLCGTLAENKRVPEVIFTSPESVMAAFEGEYFIGDDLHPSKRVRLSTKSEELVNGLVYL	897
QY	494	-----	493
Db	898	LNSLGISAIKIRFDSGYRYLVNBEELPFLGNRRKRNKAAVYSHVLPKELIETFGKQFOKMM	957
QY	494	-----	493
Db	958	SPAKLNEKVEKELDACKARRIAMLEGGDIVLDREKVTVEDEGYVYDLSVENENFLA	1017
QY	494	-----NSYGYMGYKPARRWYSKCEGASVTANGRHYTEMTKLEIKKFGKRYIAD--	543
Db	1018	GFGMLYAHNSTYIGYIYAKARWICRBCAESVTAMGNSYLETITRIEENKFGKRYVLADSV	1077
QY	544	-----	543

```

Db 1078 AGNTVEIIRNGKVEFVEPIEKLQGRVDYRIGKEKECALGVEALLTDNRGLRWKRPYI 1137
QY 544 ----- 543
Db 1138 MRHKTKKRYWFTNSWYLDVTEHSLIGLYLNTSKYKSEKPLERLVEKPRELGEKVK 1197
QY 544 ----- 543
Db 1198 SLITNRAIRSIKANDIAVRLMELIGLVGDGNGGSHKWKAYVGLSCGLDKAEIEEK 1257
QY 544 ----- 543
Db 1258 VLRLPLKAGIISNYGSKKGVDSIISKWLAGEWVKYFKDENGKRIPISEFMLPREYIE 1317
QY 544 ----- 543
Db 1318 AFLRLGFSADGIVSLRGIPIRLTSVNRSLNEVRKLLMLVGVSNMFTETTPNKYLGN 1377
QY 544 ----- 543
Db 1378 ESGTRSIHVIRKHKRFAKRIKGFELDRKATKLSNDLREHTNKKMAYDEDLVYPKIEE 1437
QY 544 ----- 543
Db 1438 INTDRYVYDIEVEGTHRFNGLVNTDGFATIPGADAEYVKKAMEFLKTIYNAKLP 1497
QY 577 LLELEVEGYFLRGFFVAKKRAVYDEGRITTTGLVRRDSEIAKEOAKYLEALIKE 636
Db 1498 LLELEVEGYFLRGFFVAKKRAVYDEGRITTTGLVRRDSEIAKEOAKYLEALIKE 1557
QY 637 DSVKAEVYKDVVEELAKVQVLEKVIYHQTITKLSKRAIGPHVIAKRLAAKIKY 696
Db 1558 GDVEEAIRIVKEVETISKYEVEPEKLVIEHQTITRDLKDVAKGPHVAVAKRLAAKIKI 1617
QY 697 RPTGIISYIVLRSGSKISDVILLSEYDPKKHXYDPDYIENOVLPVILILEAFGRKE 756
Db 1618 RPTGIISYIVLRSGSKISDVILLSEYDPKKHXYDPDYIENOVLPVILILEAFGRKE 1677
QY 757 DLYQSSKQVGLDAMLK 773
Db 1678 DLYQSSKQVGLDAMLK 1694

RESULT 13
DPOL_THERHY STANDARD; PRT; 1668 AA.
AC 09HH05;
RT 16-OCT-2001 (Rel. 40, Last Created)
RT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA polymerase (EC 2.7.7.7) [Contains: Endonuclease PI-TyII
DE (EC 3.1.-.-) (Thy pol-1 intein); Endonuclease PI-TyI (EC 3.1.-.-)
DE (Thy pol-2 intein)] (Fragment).
GN POL.
OS Thermococcus hydrothermalis.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX NCBI:Taxid=46539;
RN [1]
RN SEQUENCE FROM N.A.
RA Querellou J.J.E., Cambon M.A., Lesongeur F.O., Barbier G.:
RT "Thermococcales taxonomy and phylogeny based on the comparative use of
RT 16S rDNA, 16S-23S rDNA intergenic spacer and family B DNA polymerase
RT genes."
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RN CHARACTERIZATION OF PI-TYI.
RX MEDLINE=20512590; PubMed=11058140;
RA Saves I., Eleaume H., Dietrich J., Masson J.-M.:
RT "The Thy pol-2 intein of Thermococcus hydrothermalis is an
RT isoschizomer of PI-TyI and PI-TyII endonucleases."
RL Nucleic Acids Res. 28:4391-4396(2000).
CC -1- FUNCTION: IN ADDITION TO POLYMERASE ACTIVITY, THIS DNA POLYMERASE

```

```

CC EXHIBITS 3' TO 5' EXONUCLEASE ACTIVITY (BY SIMILARITY).
CC -1- FUNCTION: PI-TYI AND PI-TYII ARE ENDONUCLEASES. PI-TYI CLEAVES
CC THE INTERNESS SEQUENCE OF THE THY DNA POL GENE. IT REQUIRES A 21-
CC BP MINIMAL RECOGNITION SEQUENCE.
CC -1- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + (DNA)(N).
CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
CC A POST-TRANSLATIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
CC FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -1- SIMILARITY: IN THE INTEIN SECTION; BELONGS TO THE HOMING
CC ENDONUCLEASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ245819; CAC18555.1; -.
DR REBASE: 4832; PI-TyI.
DR InterPro: IPR002064; DNA_POL_B.
DR InterPro: IPR003586; HINIC.
DR InterPro: IPR003587; HININ.
DR InterPro: IPR002203; InteIn.
DR InterPro: IPR004042; InteIn_endonuc.
DR Pfam: PF00136; DNA_pol_B_3.
DR Pfam: PF03104; DNA_pol_B_exo_1.
DR PRINTS: PR00106; DNAPOLB.
DR PRINTS: PR00379; INTEIN.
DR SMART: SM00305; HINIC_2.
DR SMART: SM00306; HININ_2.
DR SMART: SM00486; POLBIC_1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; FALSE_NEG.
DR PROSITE: PS00881; PROTEIN_SPLICING_1.
DR TRANSFERASE: DNA-directed DNA polymerase; DNA replication.
KW DNA-binding; Hydrolyase; Nuclease; Exonuclease; Multifunctional enzyme;
KW Protein splicing; Autocatalytic cleavage; Endonuclease; Intron homing.
FT NON_TER 1 1
FT CHAIN 1 458 DNA POLYMERASE, 1ST PART.
FT CHAIN 459 995 ENDONUCLEASE PI-TYI (POTENTIAL).
FT CHAIN 996 1044 DNA POLYMERASE, 2ND PART.
FT CHAIN 1045 1433 ENDONUCLEASE PI-TYII (POTENTIAL).
FT CHAIN 1434 1668 DNA POLYMERASE, 3RD PART.
SQ SEQUENCE 1668 AA; 193319 MW; 5BE8805E8EDA71C8 CRC64;

Query Match 63.0%; Score 2535.5; DB 1; Length 1668;
Best Local Similarity 33.6%; Pred. No. 5e-135;
Matches 560; Conservative 85; Mismatches 92; Indels 929; Gaps 4;

QY 34 FOPYTALLKDDSAIDEIKAKGEHKGKIVYDAVKKKKFLGRGVYWKLFEEPRDY 93
Db 1 FEBYTALLKDDSAIEVKKITAGRGVYKVAERAKKFLGRPLEYWKLFETHRDY 60
QY 94 PALRGKIREHPAVIDIYEDYIPFAKRYLDKGLIPMGDEELMAFDIETFYHEDERG 153
Db 61 PAIRDEIRHSAVVDIYEDYIPFAKRYLDKGLIPMGDEELMAFDIETFYHEDERG 120
QY 154 KGEIIMISYADEEARVITWKNDLPYVDVYSNEREMIKRFOVIVREKDPVLTNGDN 213
Db 121 TGPILMISYADEEARVITWKNDLPYVEVYSTEKEIKRFLVYREKDPVLTNGDN 180
QY 214 FDLPLYIKRAEKGVTLLGGRDKEHPEPKTHRGDSFAVIRGIRHEDLPVVRRTINP 273
Db 181 FDRPVLKRCCKIGIKFTLRDGS--EPKIQRGDPRFAVEVKRIHFDLPVLRRTINP 238
QY 274 TYTLEAVYEAVLTKTSKGAEEIAAIWETEESMKKLAQYSMDATATYELGKEFPMEA 333
Db 239 TYTLEAVYEAVFTPKKRYPEBITTAWETGELEKRVARISMDATATYELGKEFPMEA 298

```



```

QY 334 ELAKLIGOSVMDVSRSTGNLVEWYLLRYAVERNELAPKPDDEERYRRLFTTYLGYYK 393
DB 299 QLSRLIGOSIMDVSRSSGNLVEWYLLRYAVERNELAPKPDDEERYRRLFTTYLGYYK 357
QY 394 EPERGLMENTYLDRCYLPSTIYTHVNSPDTLEBEGCKNVDVIVGYKFCDFPGFIP 453
DB 358 EPERLBMNIVYLDMSYPSIITHVNSPDTFNREGCKEYDTAPQVGHKFCQDVGFI 417
QY 454 SILGLIMROBIRKKMKATIDPIKKMDYRQRAVKLHA----- 493
DB 418 SILGLALBERQIKRKKASIDPLEKKLDYRQRAIKILANSLEPEMIPVENGKVRILH 477
QY 494 ----- 493
DB 478 RIGEFVDKLMTDSELVLRNGDTEVLEVRGIRALSFDKSKKARVMPYKAVIRHRYSGDV 537
QY 494 ----- 493
DB 538 YEIVLSGRRTYVTEGHSIFAYGDELNEVTGGEIKAGDILAVPRRVLPEKKERLNLVE 597
QY 494 ----- 493
DB 558 LLRLPEBETGDIILTIPYKGRKNFFKGMRLTLMISGEKKRPRTARRYLEHLEGLGYVR 657
QY 494 ----- 493
DB 658 LKKIGEYTDREGLERYKRYLRLVLEAVRYNNGKREYLVFNAVADVIALMPEELRDL 717
QY 494 ----- 493
DB 718 VGRNGFRMRPEVEIEEDFAKLLGYVSEGNARKRNOKNGMSTYKLYENQRYLDM 777
QY 494 ----- 493
DB 778 SLAEFFGVRKGRKNTIIPRKMAIIEENLCGLAENKRYPEALFTSPESYRMAFISGY 837
QY 494 ----- 493
DB 838 FIGGDVHPSKRVRLSTSELLVNLVLLNSLGVSAIKIRHDSGVYRYVNEELPFTDY 897
QY 494 ----- 493
DB 898 RKKKNAYSHVLPKEILETEFGKVFQRSVSEKPRELVSEKLDGEKAKRLEMLNGDV 957
QY 494 ----- 516
DB 958 LDKVLEVKRRPEGYVYDLSVEEDENFLAGFGLYAHNSYIGYGYARARWYCKECCASV 1017
QY 517 TAMGRHYIEMTIKEIEKFGKVLAD----- 543
DB 1018 TAMGRDIETTHIEERFGFVLADSVTGETELIIRKNGKVEFVAIEELFORVDYRIG 1077
QY 544 ----- 543
DB 1078 EKEYCVLEGEVALTLNNGRLVWKSVPYMRHRTNRKIRYVWFTNSWLYDTEHSLIGY 1137
QY 544 ----- 543
DB 1138 MNTSKVYKPKLKERLVEVKGELCESYKSLTPNRAIAHGIRVPIAVKIMELIGLVG 1197
QY 544 ----- 543
DB 1198 DGNMGGSNMAYNYGLSLGDKKEIEERILKPLKNTGIISNYDKSKKGDVSIISKMLA 1257
QY 544 ----- 543
DB 1258 RFMYVYFKDESGSKRIPEFMNLPREYIEAPLGLFSADGVSLKKGVEVRLTSVNBEL 1317
QY 544 ----- 543
DB 1318 SSSVAKLLMLVGVNSMFEVETNPNRYLCKESGTHSVHVRIDKHKRFAERIGLDRKATK 1377
QY 544 -----TDGF 547

```

```

DB 1378 LSENIGHTSKKRAKYDEDLVPPKVEIADGVYVYDVEBGTTHRFANGILVHNTDGF 1437
QY 548 YATIDGKRPETIKRKAKEFLKYNSKLPGLLEVEGYFLRGFFVAKKRYAVIDEGRIT 607
DB 1438 FATIGDAEYKVKRKAKEFLKYNAKLPGLLEVEGYFVRGFFVAKKRYAVIDEGRIT 1497
QY 608 TGLEVVRDMSKIEKQAKVLEALIKEDSVKAEVYKDVVEELAKOVPLEKLYTHE 667
DB 1498 TGLEIVVRDMSKIEKQAKVLEALIRHGVDAEAVRYKDVTEKLSKYEVEPERKLYTHE 1557
QY 668 QITKDLSEYKAIGPVAIAKRLAAGIKVRPCTIISYIYLRSGSKISDRVILLSEYDPK 727
DB 1558 QITRELKQKATGPHVIAKRLAAGIKVRPCTIISYIYLRSGSKISDRVILLSEYDPK 1617
QY 728 HKYDPDYIENOVPLAVLILFAFGYRKEDLKYQSSKOYGLDAMLK 773
DB 1618 HRYDAEYIENOVPLAVERILKAFGKKEKELRYQTRQVGLGAMLK 1663

```

```

RESULT 14
ID DPOL_METVO STANDARD: PRT: 824 AA.
AC P52025;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL.
OS Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2188;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=95014087; PubMed=7929013;
RA Konisky J., Paule S.M., Carinato M.E., Kansy J.W.;
RT "The DNA polymerase gene from the methanogenic archaeon Methanococcus
RL voltae".
CC J. Bacteriol. 176:6402-6403(1994).
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate = N diphosphate
CC + [DNA](N).
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L33366; AAA72443.1; -
DR HSSP: P56689; ITGO.
DR InterPro: IPR002064; DNA_POL_B.
DR Pfam: PF00136; DNA_POL_B.1.
DR PRINTS: PF03104; DNA_POL_B_exo; 2.
DR SMART: SM00486; POLB; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW Transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding.
SQ SEQUENCE 824 AA: 96754 MW: 094A630D8C1514FC CAC64:

```

```

Query Match 34.6%; Score 1392; DB 1; Length 824;
Best Local Similarity 36.1%; Pred. NO.4.2e-71;
Matches 308; Conservative 179; Mismatches 251; Indels 116; Gaps 21;
QY 4 DTDYITKDGKPIIRIFKKEG-----EFKIEDPHOPPIYALIKDASAI-----D 49
DB 2 DLDVNSKD--LCIDMYRNCGLKRPKEINLQKCEKFPYFYVDTSEPKIYDYLDLNOEI 59

```

```

QY 50 EIRAKGERHC---KIVRVDAVKVKKFEL-----GRDV-----EWKLI 86
DB 60 DLKLEDEFENNTSLKVQDLTNIIEIKIYSDYIINGKDISVSDFKNNKERRICKY 119
QY 87 FEHQDVPALRGKIREHPAVIDEYDIPFAKRYLIDKGLIPM-----EGDEELK 136
DB 120 VKYNNHAKIIRIYKKEG---KSEFIDIPLRKRMIDODIYPSAKYSEDKNDKINSIPELN 176
QY 137 LMAEDIEFVHEGDEFGKEGELIMISYADEBARVITKNDLP---YVDVSNEREMIK 192
DB 177 CIARDMELTYCKKEPNNAKDPIMVNLFSKDYQVITYTKKFENSNYGCYDVKDEKELIQ 236
QY 193 RFQVIVAKEDPVLITNGNFDLPYLKRAEKIGVLLGLROKHEHP---KIHMGDS 249
DB 237 KTEIL--KQYDVIYTYNGNDFPPLKKNRNIYEILDFDNASNOOPOLIKSKGIN 294
QY 250 FAVEIKRIFHDLFPVVRRTINLTPTYLEAVYEAVLIGTKSKLGAEEIAIMWTEESMKK 309
DB 295 RKSIRPILIHDLPIARKLNLTKYLENNVOLEFKINKEAVDYGDIPIKMETEDT--T 352
QY 310 LAQYMEDARATYELGKEFFPMEAEALKLGQSVYDVSRSSTGNLVEMYLRLVAYERNEL 369
DB 353 LIRAYADALYTYKMGVFLPLEIMFSRIYNOPLDYTSRMNSQOYFELLKRSFEQNM 412
QY 370 APNKPDEEYRRRLRTYLGQYVKEPBGMENTYIDFCLPSIIVTHVSPDLT--E 427
DB 413 SPNPPSSSYRERAKFSYEGYVEPLKQIEDIVSLDFMSLYPSILSHNISPEYIYE 472
QY 428 REGCKNDVAPIYGVKFCCKDFPFIPISELITMRQEIKKMKATIDPIEK-----K 480
DB 473 EKEENNEL-----GIPTKTNELLSRKHIMLK---DKIQKNFDEEYS 516
QY 481 MLDYRQAVKLANSYGYGYPKARYSKCAESVTAMGRHYIEMTIKEIEKFGKVL 540
DB 517 RLEHEOKSIVKLANSHGYLAFPARWYSDCAEMVTGLRKTYOETIEKAE--FGKVI 575
QY 541 YADPDGYATIPGKEPETIKK-----AKELKAYINKSLGGL 578
DB 576 YADPDGYAAMWDYDKLQKGGKEENDSKLSNLPKLSKEELIITTKFGLINELPEGM 635
QY 579 ELETEGYLFGFVAKKRYAVIDEGRITRTGLEVVRWDSIAKERQAVKLEAILKEDS 638
DB 636 ELEEGEFGKGLFVTKKVALIEDDGHIVKGLEVVRWDSNIAKQVAVIRALLDGD 695
QY 639 VEKAVEIVKQVVERIAKYQVPLEKVLVHEOITYDLSYKAIQPHVAIAKLAAGIYVR 698
DB 696 VNLAKKIKITKIDNLKGNIDKNDLIHTQITKNIIEYKSTAPIIEVAKKIKORGDSVR 755
QY 699 GTIISYVLGSGKISRVILLSEYDPKHHKYPDVIENOVLPVARIIEAGYRKEDL 758
DB 756 GDVISTYIVGSSRSISERAEEL--EY---ACDYDINVIYIDNOVLPVIRIMESLGSIDEL 811
QY 759 KYOSKQVGLDAML 772
DB 812 K-NSGKQFKLDQFM 824

```

```

RESULT 15
DPOL_ARCFU
ID DPOL_ARCFU STANDARD: PRT: 781 AA.
AC 029753:
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA polymerase (EC 2.7.7.7).
GN POL OR POLB OR AF0497.
OS Archaeoglobus fulgidus.
OC Archaea: Euryarchaeota: Archaeoglobales: Archaeoglobaceae:
OC Archaeoglobus.
OX NCBI_TaxID=2234:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

```

```

RX MEDLINE=98049343; PubMed=9389475;
RA Kleen H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson R.D., Kierlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.R., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Meson T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -I- CATALYTIC ACTIVITY: N deoxynucleoside triphosphate - N diphosphate
CC + (DNA)(N).
CC -I- SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001070; AAB90741.1; -
DR HSSP: P56689; ITGO.
DR TRGR: AF0497; -
DR InterPro: IPR002064; DNA_POL_B.
DR Pfam: PF00136; DNA_POL_B; 1.
DR Pfam: PF03104; DNA_POL_B_exo; 1.
DR PRINTS: PR00106; DNAPOLB.
DR SMART: SM00486; POLBc; 1.
DR PROSITE: PS00116; DNA_POLYMERASE_B; 1.
KW transferase; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Complete proteome.
SQ SEQUENCE 781 AA; 89850 MW; 761C7B18FCC61B8B CRC64;

```

```

Query Match 32.3%; Score 1301; DB 1; Length 781;
Best Local Similarity 37.4%; Pred. No. 5,2e-66;
Matches 305; Conservative 155; Mismatches 266; Indels 90; Gaps 21;
QY 2 IFPTDYITKCGKPIIRIFKKEGFEKTELDPHFOPTIYALAKDSADIEKAIGERHGK 61
DB 8 LIDADYETIGKAVVRIMCKDDGIFAYADYNDPYYIVGDE---DILKNATSTRRE 64
QY 62 IYRVYDAVKYKKFGLGADVVMKLIPEHPODVPALRGKIREHPAVIDIYDIPFAKRYL 121
DB 65 VIRKSEKQALTKTLGREGVGYIVYAHHPQHVPLRBYLSQFC---DVRADDPFAVRYL 121
QY 122 IDKGL-----IPMEGDE-----ELKIMAPDIETFFHEG--DEFG 153
DB 122 IDKGLACMGDAIEGEGQGVINSYKIEKVERIPRNEFFPLKMLVPCCEMLSSFGPEPE 181
QY 154 KGEIIMISYADEBARVITKKNIDLPYDVVSNEREMIKRFQVIVREKDPVLITNGDN 213
DB 182 KDPITIVSVKTNDDLEI-----LTGDEKTIISFVALLKSYDDPDIIVGNDA 230
QY 214 FDLPYLIRAKRAKLGVTLLGLROKHEHPKITHMGDSFAVEIKRIFHDLFPVVRRTINLP 273
DB 231 FDMPPYLRKRAERNIPLDVGRGDSN--VYFRGG---RPRITGRLLVNDLVDIAMRISDIK 284
QY 274 TYTLEAVYEAVLGTK- SKLGAEEIAIMWTEESMKKLAQYMEDARATYELGKEFFPME 332
DB 285 IKKLENNVAEFLGYKIEADIEADIDIRYWSRGEK--EKVLYARQADAINLYLAKELPPM 343
QY 333 AELAKLIGQSVMPDVSRSSTGNLVEMYLRLVAYARNELAPKRPDEEYRRRLRTYLGQYV 392
DB 344 YELSKMIRLPVDVYTRMGKQYDWMLLSEAKKIGELIAPRPHAE-----SYGARY 396

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2002, 14:17:02 ; Search time 23.89 Seconds
(without alignments)
3113.150 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026

Sequence: 1 MIFPDYITDKGKPIIRIFK.....KEDLKYSKQVGDAMLKK 774

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Maximum Match 0%

Database : PIR-71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	3356	83.4	1702 2	S42459 DNA-directed DNA P
2	3204.5	79.6	775 2	S67920 DNA-directed DNA P
3	3164.5	78.6	771 2	C75023 DNA polymerase I P
4	3151	78.3	775 2	S35343 DNA-directed DNA P
5	2934.5	72.9	1312 2	S68593 DNA-directed DNA P
6	2886	71.7	1235 2	S71210 Probable DNA-directed DNA P
7	2764.5	68.7	1670 2	S71551 DNA-directed DNA P
8	1301	32.3	781 2	A69312 DNA polymerase BI
9	1291	32.1	1634 2	E64410 DNA-directed DNA P
10	1209.5	30.0	586 2	C69028 DNA-dependent DNA P
11	1204.5	29.9	781 2	CJ7382 DNA-directed DNA P
12	1190	29.6	803 2	B56277 DNA-directed DNA P
13	1138	28.3	784 2	E72515 Probable DNA-direc
14	693.5	17.2	1088 2	T05731 DNA-directed DNA P
15	676.5	16.8	764 2	S75407 Probable DNA-direc
16	659.5	16.4	1038 1	JC5757 DNA-directed DNA P
17	657.5	16.3	1038 2	T18222 DNA polymerase del
18	657.5	16.3	1086 2	T40242 DNA polymerase del
19	649.5	16.1	1086 2	T43266 DNA-directed DNA P
20	626	15.5	901 2	E84210 DNA polymerase BI
21	621.5	15.4	1097 1	RNBRL3 DNA-directed DNA P
22	619.5	15.4	882 2	S23019 DNA-directed DNA P
23	619.5	15.4	1084 1	S19661 DNA-directed DNA P
24	616.5	15.3	872 2	JC7380 DNA-directed DNA P
25	616.5	15.3	882 2	Q90201 DNA polymerase I (
26	616.5	15.3	1105 1	S40243 DNA-directed DNA P
27	614.5	15.3	1094 2	S22573 DNA-directed DNA P
28	611	15.2	1107 1	A41618 DNA-directed DNA P
29	609	15.1	1513 2	T28158 Probable DNA-direc

30	605.5	15.0	1106 1	A39299 DNA-directed DNA P
31	587	14.6	1492 2	T18560 DNA-directed DNA P
32	584.5	14.5	879 2	A56277 DNA-directed DNA P
33	583.5	14.5	1462 1	DJH0AC DNA-directed DNA P
34	581	14.4	1081 2	T20698 hypothetical prote
35	578.5	14.4	875 2	JC5186 DNA-directed DNA P
36	570.5	14.2	959 2	F72763 Probable DNA-direc
37	570	14.2	1465 2	S45628 DNA-directed DNA P
38	559.5	13.9	223 2	E69125 DNA-dependent DNA P
39	548.5	13.6	1339 1	S20052 DNA-directed DNA P
40	527.5	13.1	1468 2	S58250 DNA-directed DNA P
41	522.5	13.0	1505 2	S28079 DNA-directed DNA P
42	516.5	12.8	844 2	T31321 DNA-directed DNA P
43	516	12.8	1405 1	DJZPA DNA-directed DNA P
44	504.5	12.5	2154 2	A84669 hypothetical prote
45	500.5	12.4	913 2	T17675 DNA-directed DNA P

ALIGNMENTS

RESULT 1

S42459

DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Thermoco

N:Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II;

C:Species: Thermococcus littoralis

C>Date: 31-Dec-1993 #sequence-revision 02-Aug-1994 #text-change 18-Jun-1999

C:Accession: S42459; S42451; S42450; S42458

R:Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, submitted to the EMBL Data Library, September 1992

A:Reference number: S42458

A:Accession: S42459

A:Molecule type: DNA

A:Residues: 1-1702 <PERI>

A:Cross-references: EMBL:M74199; NID:9154685; PIDN:AAAT2900.1; PID:9154686

R:Perler, F.B.; Comb, D.G.; Jack, W.E.; Moran, L.S.; Qiang, B.; Kucera, R.B.; Benner, Proc. Natl. Acad. Sci. U.S.A. 89, 5577-5581, 1992

A:Title: Intervening sequences in an Archaea DNA polymerase gene.

A:Reference number: S42450; MUID:92302285

A:Accession: S42451

A:Molecule type: DNA

A:Residues: 181-922,387-425,452-476,483-524,1021-1062,1076-1099,1466-1489,1533-1547 <A:Cross-references: EMBL:M74198

R:Hodges, R.A.; Perler, F.B.; Noren, C.J.; Jack, W.E. Nucleic Acids Res. 20, 6153-6157, 1992

A:Title: Protein splicing removes intervening sequences in an archaea DNA polymerase.

A:Reference number: S40788; MUID:93117083

A:Accession: S40788

A:Contents: annotation

C:Function: <VENT>

A:Description: nucleotidyltransferase

A>Note: DNA-directed DNA polymerase Vent

C:Function: <END1>

A:Description: endonuclease; hydrolase

A>Note: DNA endonuclease PI-TII

C:Function: <END2>

A:Description: endonuclease; hydrolase

A>Note: DNA endonuclease PI-TIII

C:Superfamily: Thermococcus littoralis DNA-directed DNA polymerase Vent

C:Keywords: DNA binding; endonuclease; hydrolase; nucleotidyltransferase; protein spl

F:1-494/Domain: DNA-directed DNA polymerase Vent extein 1 #status predicted <XT1>

F:495-1032/Product: DNA endonuclease PI-TI (pol Vent extein 1) #status predicted <MAT

F:1033-1081/Domain: DNA-directed DNA polymerase Vent extein 2 #status predicted <MAT

F:1082-1471/Product: DNA endonuclease PI-I (pol Vent extein 2) #status predicted <MAT

F:1472-1702/Domain: DNA-directed DNA polymerase Vent extein 3 #status predicted <XT3>

F:494-1033/Cross-Link: peptide (Asn-Ser) #status predicted

F:1081-1472/Cross-Link: peptide (Asp-Thr) #status predicted

Query Match 83.4% Score 3356; DB 2; Length 1702;
Best Local Similarity 42.7% Pred. NO. 1,2e-164;
Matches 726; Conservative 33; Mismatches 15; Indels 928; Gaps 2;

```

QY 1 MIFDQDITTKGKPKIRIRFKKENGFEKIEDPHQPIYIALLKDSADDEIKAIGERHG 60
Db 1 MILDQDITTKGKPKIRIRFKKENGFEKIEDPHQPIYIALLKDSADDEIKAIGERHG 60
QY 61 KIVRVDAVVKYKKRKLGDVEWVKLIFEHPOVPALRGKIREHPAVIDIEYDIPAKRY 120
Db 61 KIVRVDAVVKYKKRKLGDVEWVKLIFEHPOVPALRGKIREHPAVIDIEYDIPAKRY 120
QY 121 LIDKGLIMEGDEBELKLAFFDIEFFYHGDGEGKEIIMISYADEEARVITWKNDLPY 180
Db 121 LIDKGLIMEGDEBELKLAFFDIEFFYHGDGEGKEIIMISYADEEARVITWKNDLPY 180
QY 181 VDVYSNEEMAKRKYQVIREKDPVLTINQDNBDPLYLIRAKELGYTLILGRKHEPE 240
Db 181 VDVYSNEEMAKRKYQVIREKDPVLTINQDNBDPLYLIRAKELGYTLILGRKHEPE 240
QY 241 PKIRMGSPFAVEIKGRHFDLPFVVRRTINLPYTLAEVAVGKRSKAGAEIAI 300
Db 241 PKIRMGSPFAVEIKGRHFDLPFVVRRTINLPYTLAEVAVGKRSKAGAEIAI 300
QY 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEIAKLIGOSVMDVSSSGNLVEMYLL 360
Db 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEIAKLIGOSVMDVSSSGNLVEMYLL 360
QY 361 RVAYERNLAPNKPDEEYRRRLRTTYIGYVKEPERGLMENTITLDRCLYPSITVHN 420
Db 361 RVAYERNLAPNKPDEEYRRRLRTTYIGYVKEPERGLMENTITLDRCLYPSITVHN 420
QY 421 VSPDLEEGCKNDVAVIVGKFCCKDPGFIPISTLIGLITWROEIKKKMATIPIEKK 480
Db 421 VSPDLEEGCKNDVAVIVGKFCCKDPGFIPISTLIGLITWROEIKKKMATIPIEKK 480
QY 481 MLDYRQRAVKAHA----- 493
Db 481 MLDYRQRAVKAHA----- 493
QY 494 ----- 493
Db 494 ----- 493
QY 541 VNNLPASFNNKIKESYKVKYKALIRHKYKKAWEIQLSSGRKINITAGHSLFTVNGEI 600
Db 541 VNNLPASFNNKIKESYKVKYKALIRHKYKKAWEIQLSSGRKINITAGHSLFTVNGEI 600
QY 601 KEVSGDGIEKEDLIVAPKKIKLNEKVSINIPELISDSEETADIVMTISAKGKNFK 660
Db 601 KEVSGDGIEKEDLIVAPKKIKLNEKVSINIPELISDSEETADIVMTISAKGKNFK 660
QY 661 GMLRTLRMGEGENRRRTFNRYLFHLEKGLIKLPRGYEVDWERLKKYKQLEYKLAG 720
Db 661 GMLRTLRMGEGENRRRTFNRYLFHLEKGLIKLPRGYEVDWERLKKYKQLEYKLAG 720
QY 721 SVKINGNKRRELVMPNEIKDFISYPOKLEEMWIKITLNGFRINCILAVDEDFKLLGY 780
Db 721 SVKINGNKRRELVMPNEIKDFISYPOKLEEMWIKITLNGFRINCILAVDEDFKLLGY 780
QY 781 VSEGYAGAKKKTGISTSVKLYNEDPNVLESMMKNAKFGKAVDNCVYSISKMAVL 840
Db 781 VSEGYAGAKKKTGISTSVKLYNEDPNVLESMMKNAKFGKAVDNCVYSISKMAVL 840
QY 841 VMKCLGALAEKRIPIVILTSPEVRWMSLEAVFTGDGDIIHPSKRRLSTKSELLANOL 900
Db 841 VMKCLGALAEKRIPIVILTSPEVRWMSLEAVFTGDGDIIHPSKRRLSTKSELLANOL 900
QY 901 VFLNLSLGISSVKIGFDSGVRYVINELOPOTSREKNTYYSNLIPKEILLRDVFGKEFO 960
Db 901 VFLNLSLGISSVKIGFDSGVRYVINELOPOTSREKNTYYSNLIPKEILLRDVFGKEFO 960
QY 961 KMMFFKKRELVDGKLNREKAKLLEFINGDVLDRVSKVKEKQYDEGVYVLDSEVDEN 1020
Db 961 KMMFFKKRELVDGKLNREKAKLLEFINGDVLDRVSKVKEKQYDEGVYVLDSEVDEN 1020
QY 1021 FLVFGGLYAHNSYGYGMYGPKARWYSKECAESYTAGMRHYIEMTIREIEEKGFGKVLYA 1080
Db 1021 FLVFGGLYAHNSYGYGMYGPKARWYSKECAESYTAGMRHYIEMTIREIEEKGFGKVLYA 1080
QY 543 D----- 543

```

```

Db 1081 DSVGSESEIIRQNGKIRFVKIKDLFSKVYSIGERKEYCILGVEALTLDDGKLWMPV 1140
QY 544 ----- 543
Db 1141 PYVMHRANKRMFRIMWLNSTWIDVTEHSLIGYLTNTSKTKAKKIGERLKEVKKPEELG 1200
QY 544 ----- 543
Db 1201 AVKSLICPNAPLKDENTKTSIAVAFWELVGLIVDGNMGDSRNAEYVIGLSTGDAE 1260
QY 544 ----- 543
Db 1261 IKQKLEPLKTYGVSINYYPKNEKGFNLAKSLVKFMKRHEKDEKRRKIPFMYELPV 1320
QY 544 ----- 543
Db 1321 TYIEAFNLGFSADGVIRKGVPEIRLITNADFLREVRKLLMIVGISNSIFAETTPNR 1380
QY 544 ----- 543
Db 1381 YNGVSTGYSKHLRIKKNRFAERIGFLIERKQRLLEHLSARVRNRTIDGFDLVHVK 1440
QY 544 ----- 543
Db 1441 KVEIIPYGGYVYDIEVETHEFPANNILVHNDGFYATIPGKPELIRKKAKEFLYINS 1500
QY 573 KLPGLELEEGFYLDGFFVAKKRYAVIDEGRITTRGLVVRDMSIAKETQAKVLEA 632
Db 1501 KLPGLELEEGFYLDGFFVAKKRYAVIDEGRITTRGLVVRDMSIAKETQAKVLEA 632
QY 633 ILKEDSVKAEIVADVBEIAKYOVPLEKLYIHQITKOLSEYKAIQPHVAIAKRLAK 692
Db 1561 ILKEDSVKAEIVADVBEIAKYOVPLEKLYIHQITKOLSEYKAIQPHVAIAKRLAK 692
QY 693 GIKVRPGTIIISYIVRGSGKISDRVITLSEYDPRKHKTDPOVYIENOVLPAYRLTEAFG 752
Db 1621 GIKVRPGTIIISYIVRGSGKISDRVITLSEYDPRKHKTDPOVYIENOVLPAYRLTEAFG 752
QY 753 YRKEDLKYOSSKQVGLDAMLR 774
Db 1681 YRKEDLKYOSSKQVGLDAMLR 774

```

RESULT 2
 S67920
 DNA-directed DNA polymerase (EC 2.7.7.7) - Thermococcus sp.
 C:Species: Thermococcus sp.
 C:Date: 17-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 18-Jun-1999
 C:Accession: S67920
 R:Southworth, M.S.; Kong, H.; Kucera, R.B.; Ware, J.; Jannasch, H.W.; Perler, F.B.
 submitted to the EMBL Data Library, January 1996
 A:Description: Cloning, expression and modulation of the 3'-5' exonuclease activity
 A:Reference number: S67920
 A:Accession: S67920
 A:Molecule type: DNA
 A:Residues: 1-775 <SOU>
 A:Cross-references: GB:U47108; NID:g1197451; PIDN:AAA86769.1; PID:g1197452
 A:Experimental source: strain 90N-7
 C:Superfamily: herpesvirus DNA-directed DNA polymerase
 C:Keywords: DNA binding; nucleotidyltransferase

Query Match 79.6%, Score 3204.5; DB 2; Length 775;
 Best Local Similarity 76.6%, Pred. No. 2.5e-157;
 Matches 592; Conservative 89; Mismatches 89; Indels 3; Gaps 2;

```

QY 1 MIFDQDITTKGKPKIRIRFKKENGFEKIEDPHQPIYIALLKDSADDEIKAIGERHG 60
Db 1 MILDQDITTKGKPKIRIRFKKENGFEKIEDPHQPIYIALLKDSADDEIKAIGERHG 60
QY 61 KIVRVDAVVKYKKRKLGDVEWVKLIFEHPOVPALRGKIREHPAVIDIEYDIPAKRY 120
Db 61 KIVRVDAVVKYKKRKLGDVEWVKLIFEHPOVPALRGKIREHPAVIDIEYDIPAKRY 120

```

Dh 61 TVVKKRAEKYOKKFLGRPIEWKLYNHPDPAIRDRIRAHPAVVDIYEYDIPAKRY 120
Qy 121 LIDGLIPMEGDEELKMAFDIETFEHGEDEFGKEIIMISYADEEARVITWKKIDLPY 180
Dh 121 LIDGLIPMEGDEELKMAFDIETFEHGEDEFGKEIIMISYADEEARVITWKKIDLPY 180
Qy 181 VDVAVSNEREMIKRFOVQIAREKDPVLTITNGDNFDLPYLIRAKELGVTLLGRDKEHE 240
Dh 181 VDVAVSNEREMIKRFOVQIAREKDPVLTITNGDNFDLPYLIRAKELGVTLLGRDKEHE 240
Qy 241 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPYTTLEAVYEAVLGKTSKLGAEETAI 300
Dh 239 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPYTTLEAVYEAVLGKTSKLGAEETAI 300
Qy 301 WETESMKRLAQSMEBARATYELGKEFPFMEAEIATLIGOSVMDVSRSSSTGNLVEWYLL 360
Dh 299 WETESMKRLAQSMEBARATYELGKEFPFMEAEIATLIGOSVMDVSRSSSTGNLVEWYLL 360
Qy 361 RYAVERNELAPNKDEEYRRLRTTYLGGYVKEPERGLMNTIYLDRCILYPSIITVHN 420
Dh 359 RYAVERNELAPNKDEEYRRLRTTYLGGYVKEPERGLMNTIYLDRCILYPSIITVHN 420
Qy 421 VSPDTLEREGCKNTDPAIVGYKFCCKDPFGFIPSLIGLITMRQEIKKKKATIDPLEKK 480
Dh 418 VSPDTLEREGCKNTDPAIVGYKFCCKDPFGFIPSLIGLITMRQEIKKKKATIDPLEKK 480
Qy 481 MLDYRORAVKLHANSYGYMGYPKARWYKSCAEVSATAMGRHYEMTITKEIEKGFVYL 540
Dh 478 MLDYRORAVKLHANSYGYMGYPKARWYKSCAEVSATAMGRHYEMTITKEIEKGFVYL 540
Qy 541 YADTDGFYATIPGEPETIKKAKAEFLKYINSKLPGLLEEGFYLRGFFVAKKRYAVI 600
Dh 538 YADTDGFYATIPGEPETIKKAKAEFLKYINSKLPGLLEEGFYLRGFFVAKKRYAVI 600
Qy 601 DEGRITRGLGVRRMSSEIAKETOAKVLEAILKEDSVKAEVAVIVDVEIAKYOVL 660
Dh 598 DEGRITRGLGVRRMSSEIAKETOAKVLEAILKEDSVKAEVAVIVDVEIAKYOVL 660
Qy 661 EKLVIHQITRDLSEYKAIQPHVAIAKRLAKGIVRPGTIIISYIVLRGSGKISDRVILL 720
Dh 658 EKLVIHQITRDLSEYKAIQPHVAIAKRLAKGIVRPGTIIISYIVLRGSGKISDRVILL 720
Qy 721 SEYDPKKHKYDPDYIENQVLPVLRILEAFYKREKEDIKYSSKOVGLDAMLK 773
Dh 718 SEYDPKKHKYDPDYIENQVLPVLRILEAFYKREKEDIKYSSKOVGLDAMLK 773

RESULT 3
C75023
DNA polymerase I PAB1128 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C:Accession: C75023
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A:Reference number: A75001
A:Accession: C75023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-771 <RAW>
A:Cross-references: GB:AJ248288; GB:AL096836; NID:95458960; PIDN:CAB50625.1; PID:9545913
A:Experimental source: strain Orsay
C:Genetics: pol
A:Gene: pol1
C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 78.6%; Score 3164.5; DB 2; Length 771;
Best Local Similarity 76.1%; Pred. No. 2.8e-155;
Matches 588; Conservative 89; Mismatches 93; Indels 3; Gaps 2;

Qy 1 MIFDTYITKDGKPIIRIFKKEGGEFKIELDPHFQPIYIALKDSALDEIKAIKERHG 60

Qy 1 MIFDTYITKDGKPIIRIFKKEGGEFKIELDPHFQPIYIALKDSALDEIKAIKERHG 60

Dh 1 MIFDTYITKDGKPIIRIFKKEGGEFKIELDPHFQPIYIALKDSALDEIKAIKERHG 60
Qy 61 KIYAVVADAVKAKKFLGADVWVKLIFEHPODVPAIRKRIEHPAVIDIYEYDIPAKRY 120
Dh 61 KIYAVVADAVKAKKFLGADVWVKLIFEHPODVPAIRKRIEHPAVIDIYEYDIPAKRY 120
Qy 121 LIDGLIPMEGDEELKMAFDIETFEHGEDEFGKEIIMISYADEEARVITWKKIDLPY 180
Dh 121 LIDGLIPMEGDEELKMAFDIETFEHGEDEFGKEIIMISYADEEARVITWKKIDLPY 180
Qy 181 VDVAVSNEREMIKRFOVQIAREKDPVLTITNGDNFDLPYLIRAKELGVTLLGRDKEHE 240
Dh 181 VDVAVSNEREMIKRFOVQIAREKDPVLTITNGDNFDLPYLIRAKELGVTLLGRDKEHE 240
Qy 241 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPYTTLEAVYEAVLGKTSKLGAEETAI 300
Dh 239 PKIHRMGDSFAVEIKGRHFDLPFVVRTINLPYTTLEAVYEAVLGKTSKLGAEETAI 300
Qy 301 WETESMKRLAQSMEBARATYELGKEFPFMEAEIATLIGOSVMDVSRSSSTGNLVEWYLL 360
Dh 299 WETESMKRLAQSMEBARATYELGKEFPFMEAEIATLIGOSVMDVSRSSSTGNLVEWYLL 360
Qy 361 RYAVERNELAPNKDEEYRRLRTTYLGGYVKEPERGLMNTIYLDRCILYPSIITVHN 420
Dh 359 RYAVERNELAPNKDEEYRRLRTTYLGGYVKEPERGLMNTIYLDRCILYPSIITVHN 420
Qy 421 VSPDTLEREGCKNTDPAIVGYKFCCKDPFGFIPSLIGLITMRQEIKKKKATIDPLEKK 480
Dh 418 VSPDTLEREGCKNTDPAIVGYKFCCKDPFGFIPSLIGLITMRQEIKKKKATIDPLEKK 480
Qy 481 MLDYRORAVKLHANSYGYMGYPKARWYKSCAEVSATAMGRHYEMTITKEIEKGFVYL 540
Dh 478 MLDYRORAVKLHANSYGYMGYPKARWYKSCAEVSATAMGRHYEMTITKEIEKGFVYL 540
Qy 541 YADTDGFYATIPGEPETIKKAKAEFLKYINSKLPGLLEEGFYLRGFFVAKKRYAVI 600
Dh 538 YADTDGFYATIPGEPETIKKAKAEFLKYINSKLPGLLEEGFYLRGFFVAKKRYAVI 600
Qy 601 DEGRITRGLGVRRMSSEIAKETOAKVLEAILKEDSVKAEVAVIVDVEIAKYOVL 660
Dh 598 DEGRITRGLGVRRMSSEIAKETOAKVLEAILKEDSVKAEVAVIVDVEIAKYOVL 660
Qy 661 EKLVIHQITRDLSEYKAIQPHVAIAKRLAKGIVRPGTIIISYIVLRGSGKISDRVILL 720
Dh 658 EKLVIHQITRDLSEYKAIQPHVAIAKRLAKGIVRPGTIIISYIVLRGSGKISDRVILL 720
Qy 721 SEYDPKKHKYDPDYIENQVLPVLRILEAFYKREKEDIKYSSKOVGLDAMLK 773
Dh 718 SEYDPKKHKYDPDYIENQVLPVLRILEAFYKREKEDIKYSSKOVGLDAMLK 773

RESULT 4
S35543
DNA-directed DNA polymerase (EC 2.7.7.7) - Pyrococcus furiosus
C:Species: Pyrococcus furiosus
C:Date: 31-Dec-1993 #sequence_revision 02-Aug-1994 #text_change 20-Jun-2000
C:Accession: S35543
R:Umori, T.; Ishino, Y.; Toh, H.; Asada, K.; Kato, I.
Nucleic Acids Res. 21, 259-265, 1993
A:Title: Organization and nucleotide sequence of the DNA polymerase gene from the arc
A:Reference number: S35543; MUID:93181200
A:Accession: S35543
A:Molecule type: DNA
A:Residues: 1-775 <UEM1>
A:Cross-references: EMBL:ID12983; NID:9216917; PIDN:BA002362.1; PID:9216918
A:Accession: S44596
A:Molecule type: protein
A:Residues: 1-12 <UEM2>
C:Genetics: pol
A:Gene: pol
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: DNA binding; nucleotidyltransferase

C:Keywords: DNA binding; nucleotidyltransferase

F:1-775/Product: DNA-directed DNA polymerase #status experimental <MNT>

Query Match 78.3%; Score 3151; DB 2; Length 775;

Best Local Similarity 74.6%; Pred. No. 1.4e-154; Mismatches 576; Conservative 100; Matches 94; Indels 2; Gaps 1;

QY 1 MIFDDYITKDKSPITRIKKEGKGEKIELDPHFQPIYALALKDSALDEIKAIKERG 60
 1 MILDVDTIEBCKPVIRLFKKENGKFKIEHDTFRYIYALLRDSKIEVKKIGERG 60
 DB 1 MILDVDTIEBCKPVIRLFKKENGKFKIEHDTFRYIYALLRDSKIEVKKIGERG 60
 QY 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 DB 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 QY 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 DB 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 QY 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 DB 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 QY 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 DB 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 QY 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 DB 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 QY 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 DB 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 QY 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 DB 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 QY 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 DB 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 QY 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 DB 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 QY 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 DB 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 QY 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 DB 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 QY 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 DB 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 QY 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 DB 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 QY 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 DB 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 QY 541 YADTGGAFTATGEGKPEPTIKKAKKELKINKLGLLELEEGYLRGFFPAKKRYAVI 600
 541 YADTGGAFTATGEGKPEPTIKKAKKELKINKLGLLELEEGYLRGFFPAKKRYAVI 600
 DB 541 YADTGGAFTATGEGKPEPTIKKAKKELKINKLGLLELEEGYLRGFFPAKKRYAVI 600
 QY 539 YIDTGLATYIGGSESEIKKALEFVKYKINSKLPGLLELEEGYLRGFFPAKKRYAVI 598
 539 YIDTGLATYIGGSESEIKKALEFVKYKINSKLPGLLELEEGYLRGFFPAKKRYAVI 598
 DB 539 YIDTGLATYIGGSESEIKKALEFVKYKINSKLPGLLELEEGYLRGFFPAKKRYAVI 598
 QY 601 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 660
 601 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 660
 DB 601 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 660
 QY 599 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 658
 599 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 658
 DB 599 DEGRITTRGLENVRDSEIKETQAKVLEALIKEDSVEKAVEIKVDVEELAKYQVPL 658
 QY 661 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 720
 661 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 720
 DB 661 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 720
 QY 659 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 718
 659 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 718
 DB 659 EKLVIHEQITKDLSEYKALGPVAVIAIKRLAAGIKVAPGTIIISYLVNLSGKISDRVILL 718
 QY 721 SEYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 772
 721 SEYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 772
 DB 721 SEYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 772
 QY 719 EBYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 770
 719 EBYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 770
 DB 719 EBYDRKAKKHYDDYIYENOVPLAVLRILEAFGYRKEDLKYOSKQVGLDAML 770

RESULT 5

S68593 DNA-directed DNA polymerase (EC 2.7.7.7) Vent, intein containing precursor - Pyrococcus

N:Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA-directed DNA polymerase (EC 2.7.7.7)

C:Species: Pyrococcus sp.

C:Date: 24-Aug-1996 #sequence_revision 01-Nov-1996 #text_change 24-Sep-1999

C:Accession: S68593

R:Yu, M.O.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perler, F.B.

Submitted to the EMBL Data Library, August 1993

A:Description: In vitro protein splicing of purified precursor and the identification of

A:Reference number: S68593

A:Accession: S68593

A:Molecule type: DNA

A:Residues: 1-1312 <XUA>

A:Cross-references: EMBL:000707; NID:9436492; PIDN:AAA67130.1; PID:9825735

R:Yu, M.O.; Southworth, M.W.; Mersha, F.B.; Hornstra, L.J.; Perler, F.B.

Cell 75, 1371-1377, 1993

A:Title: In vitro protein splicing of purified precursor and the identification of a

A:Reference number: S68581; MUID:94094530

A:Contents: annotation

R:Yu, M.O.; Comb, D.G.; Paulus, H.; Noren, C.J.; Shao, Y.; Perler, F.B.

EMBO J. 13, 5517-5522, 1994

A:Title: Protein splicing: an analysis of the branched intermediate and its resolution

A:Reference number: S52065; MUID:95080235

A:Contents: annotation; self-splicing mechanism

A:Description: nucleotidyltransferase

A:Note: DNA-directed DNA polymerase Vent

A:Function: <ENDO>

A:Function: endonuclease; hydrolase

A:Note: DNA endonuclease PI-1

A:Superfamily: hypothetical protein PH0202

C:Keywords: DNA replication; endonuclease; hydrolase; nucleotidyltransferase; protein

F:1-491,1030-1312/Product: DNA-directed DNA polymerase Vent #status predicted <MNT>

F:1-491/DNA: DNA-directed DNA polymerase Vent extein 1 #status predicted <MNT>

F:492-1029/Product: DNA endonuclease PI-1 (pol Vent intein 1) #status predicted <MNT>

F:1030-1312/DNA: DNA endonuclease PI-1 (pol Vent intein 2) #status predicted <MNT>

F:491-1030/Cross-link: peptide (Ala-Ser) #status experimental

Query Match

Best Local Similarity 72.9%; Score 2934.5; DB 2; Length 1312;

Matches 597; Conservative 86; Mismatches 87; Indels 539; Gaps 3;

QY 1 MIFDDYITKDKSPITRIKKEGKGEKIELDPHFQPIYALALKDSALDEIKAIKERG 60
 1 MILDVDTIEBCKPVIRLFKKENGKFKIEHDTFRYIYALLRDSKIEVKKIGERG 60
 DB 1 MILDVDTIEBCKPVIRLFKKENGKFKIEHDTFRYIYALLRDSKIEVKKIGERG 60
 QY 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 DB 61 KIVRVAVDAKVKKKFLGRDVEWKKLFEHPDVPALRGIRHPAVIDIYEIDIPAKRY 120
 QY 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 DB 121 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 QY 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 DB 181 VDVASNEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 240
 QY 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 DB 181 VEVSSEREMIKRFVQIVAREKDPDLITYNGDNFPLYLKRAEKLGVTLGRDKEHE 238
 QY 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 DB 241 PRHMGDSFAVEIKRIFHDLFPVVRRTINPTTYLLEAVYEAVALGKTSKGAEEIAI 300
 QY 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 DB 239 PKMQRIGDMTAEVGRHFDLYHVTITINLPTTYLLEAVYEAVALGKTSKGAEEIAI 298
 QY 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 DB 301 WETESMKKLAQYSMEDARATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 360
 QY 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 DB 299 WESGLENLEAVAKYSMEDAKATYELGKEFPMEAEIAGLIGQSVMDSSSTGNLVEWYLL 358
 QY 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 DB 361 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 420
 QY 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 DB 359 RYAVERNELAPKPDDEEYRRRLRTTYLGGYKKEPBGIMENITYIDFRCLYPSIIYTHN 418
 QY 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 DB 421 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 480
 QY 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 DB 419 VSPDITLEREGCKNVDVAPVIGYKFCDFPGLFIPSLIGELITMROEIKKKMKATIDPIERK 478
 QY 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 DB 481 MLDYRQRAIKILANSILPDEWPLIKNGKYKIPRIDFVGLMKANQGVKKTGTDEYLE 538
 QY 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 DB 479 LIDKGLIPMEGDEELKILAFDIETLYHGEDEFGKPIIMISADEBEAKVITWKNDLPY 180
 QY 491 ---LNA----- 493
 491 ---LNA----- 493
 DB 491 ---LNA----- 493
 QY 539 VAGIHAFSPDRSKKARVAVAVIRHRYSGVNYRIVLNSGKRTITTEGHSLEFVYRNGDL 598
 539 VAGIHAFSPDRSKKARVAVAVIRHRYSGVNYRIVLNSGKRTITTEGHSLEFVYRNGDL 598
 DB 539 VAGIHAFSPDRSKKARVAVAVIRHRYSGVNYRIVLNSGKRTITTEGHSLEFVYRNGDL 598
 QY 494 ----- 493

QY 620 EIAKETQAKVLEAIKEDSEKAVEIKVDVEIAKQVPLEKLVIEHQITKDLSEYKAI 679
 DB 1078 EIAKETAQVLEAIKEDSEKAVEIKVDVEIAKQVPLEKLVIEHQITKDLSEYKAI 1137
 QY 680 GHVAVAKRLAKGKIKVREGTIIISYIVLKGSKISDRVLLSEYDPKKKYPDPYIENQ 739
 DB 1138 GHVAVAKRLAKGKIKVREGTIIISYIVLKGSKISDRVLLSEYDPKKKYPDPYIENQ 1197
 QY 740 VPAVLRIIEAFGRKEDKLYOSKOVGLDAMLK 773
 DB 1198 VPAVLRIIEAFGRKEDKLYOSKOVGLDAMLK 1231

RESULT 7
 S71551
 DNA-directed DNA polymerase (EC 2.7.7.7) KOD, intein containing precursor - *Pyrococcus* sp.
 N:contants: DNA endonuclease (EC 3.1.1.-) PI-I; DNA endonuclease (EC 3.1.1.-) PI-II; DNA
 C:Species: *Pyrococcus* sp.
 A:Variety: Strain KOD1
 C:Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Apr-2000
 C:Accession: S71551
 A:Kakihara, H.; Takagi, M.; Imanaka, T.
 A:Description: Cloning and analysis of the DNA polymerase gene from a new hyperthermophilic
 A:Reference number: S71551
 A:Accession: S71551
 A:Molecule type: DNA
 A:Residues: 1-1670 <KAK>
 C:Cross-references: EMBL:D29671
 C:Function: <Nucle>
 A:Description: as DNA-directed DNA polymerase, catalyzes the polymerization of DNA at th
 A>Note: DNA-directed DNA polymerase KOD
 C:Function: <EN1>
 A:Description: as DNA endonuclease PI-I, catalyzes the hydrolysis of internal phosphodi
 A>Note: DNA endonuclease PI-PspI
 C:Function: <EN2>
 A:Description: as DNA endonuclease PI-II, catalyzes the hydrolysis of internal phosphodi
 A>Note: DNA endonuclease PI-PspII
 C:Superfamily: DNA-directed DNA polymerase KOD
 C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing
 F:1-406/767-851/1388-1670/Product: DNA-directed DNA polymerase KOD #status predicted <MA
 F:1-406/Domain: DNA-directed DNA polymerase KOD extein 1 #status predicted <XT1>
 F:407-766/Product: DNA endonuclease PI-I (pol KOD extein 1) #status predicted <MAT2>
 F:767-851/Domain: DNA-directed DNA polymerase KOD extein 2 #status predicted <XT2>
 F:852-1387/Product: DNA endonuclease PI-II (pol KOD extein 2) #status predicted <MAT3>
 F:1388-1670/Domain: DNA-directed DNA polymerase KOD extein 3 #status predicted <XT3>
 F:406-767/Cross-link: peptide (Arg-Ser) #status predicted
 F:851-1388/Cross-link: peptide (Asn-Ser) #status predicted

Query Match 68.7%; Score 2764.5; DB 2; Length 1670;
 Best Local Similarity 35.9%; Pred. No. 2.9e-134;
 Matches 599; Conservative 83; Mismatches 88; Indels 899; Gaps 4;

QY 1 MFDDYITKDKKPIIRIFKKENGGEKIEIDPHFOPIYAILLKDSADIDEIKANGERH 60
 DB 1 MLDDYITEDKPVIRIFKKENGGEKIEYDRTFEYFALLKDSADIEVKKITAEHRG 60
 QY 61 KIVRVDAVAVKVKKGLGADVEMKLFIEHPDVPALRGKIREHPAVIDIEYEDIPAKRY 120
 DB 61 TVYTVKRVKVKKGLGADVEMKLFIEHPDVPALRGKIREHPAVIDIEYEDIPAKRY 120
 QY 121 LIDKGLIPMEGDELKLMFIDIEYFHEGDEFGKEIIMISYADEEGARVITMKNDLPY 180
 DB 121 LIDKGLIPMEGDELKLMFIDIEYFHEGDEFGKEIIMISYADEEGARVITMKNDLPY 180
 QY 181 VDVVSNRERMIKRFVQIVAEKDPVLITNGDNFPLIKRAEKIGVTLGLGRDKEHE 240
 DB 181 VDVVSNRERMIKRFVQIVAEKDPVLITNGDNFPLIKRAEKIGVTLGLGRDKEHE 240
 QY 241 PIHHRGDFAEVIEKRIHFDLPVVRTINLPTYLEAVYAVYLGKTKGAEGIAI 300
 DB 239 PIQRMGRFAVEVKGRIHFDLPVVRTINLPTYLEAVYAVYLGKTKGAEGIAI 298

QY 301 WETESMKLIAQSYMEDARATYELGKEFPMEAEIAKIGOSYWDYSRSSTGLNLYEYLL 360
 DB 299 WETESMKLIAQSYMEDARATYELGKEFPMEAEIAKIGOSYWDYSRSSTGLNLYEYLL 358
 QY 361 RYAVERNELAPKPDDEEYRRLRTTYLGGYKKEPERGLMENTYUDFRC----- 410
 DB 359 RYAVERNELAPKPDDEEYRRLRTTYLGGYKKEPERGLMENTYUDFRC----- 410
 QY 411 ----- 410
 DB 418 GGIINISEVQGDYVLIDGQVRKRWEDYKGLVNLINGLCTPNHKLPLYTKNEBO 477
 QY 411 ----- 410
 DB 478 TRIROSLAKSFLTKKVKKGIITPTPLFEIGRATSENIPEEVLKGLAGILLAEGLLRK 537
 QY 411 ----- 410
 DB 538 DVEYDSRRKKRISHQVVEITIGKDEEFDRITTYIFERLGTIPSISEKGTNAVTL 597
 QY 411 ----- 410
 DB 598 KVAKKNVLYKVEIMDNISLAPSVLRGFEEDGSVNRVRSYATGCTKMEKIKLYS 657
 QY 411 ----- 410
 DB 658 KLSQIGIPHQTYTYQYQENGKDRSRYLEITGKDLILFQTLGFISSKRNALLKAI 717
 QY 411 -----LPSIIVTN 420
 DB 718 QRENNLNNNGYRLSEFNSTVEYEGKYDILTCTPYFANGILTNHSLYPSIITN 777
 QY 421 VSPDTEREGCKNYDVAPIVGYKFCDFPGFIPSIIGELITMROEIKKKAKATIDIEKK 480
 DB 778 VSPDTEREGCKNYDVAPIVGYKFCDFPGFIPSIIGELITMROEIKKKAKATIDIEKK 837
 QY 481 MDYRORAVKLAH----- 493
 DB 838 LLDYRORAKITLANSILPEEWLPVLEEGVHFEVRIEGLIDRMMEENAGVKREGTEVLE 897
 QY 494 ----- 493
 DB 898 VSGLEVPSENRKTKAEKRVKALIHNDYSGKYTTIRLKSGRIRIKTTSCHSLFVSNGEL 957
 QY 494 ----- 493
 DB 958 VEYTGDELKPGDLVAVPRLELPERNHVNLNLELLGTPDEETLDIVMTIPVKKKNFKP 1017
 QY 494 ----- 493
 DB 1018 GMLRTLMIFFGEKRPRTARYLRHLLEDLGVRLLKIGYEVLDMSLKNYRLYEALVEN 1077
 QY 494 ----- 493
 DB 1078 VRRNGKREYLVFNISIRAVGIMPLKELKEMKIGTLNGFRKRLIEVDESLAKLIGYV 1137
 QY 494 ----- 493
 DB 1138 SSGYAKQKNPKNGMSYKLVNEDPEVLDMERLASREFGVRGRNRYVEIPKIGYLL 1197
 QY 494 ----- 493
 DB 1198 FENMCGVLAENKRIPFVETSPKGVRLAELEGYSAMATSTEOALNEKRALANQLVL 1257
 QY 494 ----- 493
 DB 1258 LINSVGSVAVKLGHSQVRYVYINELPVLKLDKKNNAVYSHVIRKVELSEVFGKVFQKN 1317
 QY 494 ----- 493
 DB 1318 VSPQTFRKVEDGRDPEKAQRLSWLIEGDYVLDRVESYVDVEDYDGYVYDLSVEDNENFL 1377

OY 494 -----NSYGYMGYPKARWYKSCAESVTAMGRHYIMTKEIEKFGFVLVADT 544
 DB 1378 VGFGLVYAHNSYGYGYARARWYCKEACSVTAMGREYIMTKEIEKFGFVLVADT 1437
 OY 545 DGFATIPGEPEPTIKKAKKFLKYINSKPLLELEEGYLGFGFYAKRVAVIDEG 604
 DB 1438 DGFATIPGEPEPTIKKAKKFLKYINSKPLLELEEGYLGFGFYAKRVAVIDEG 1497
 OY 605 RITRGLSEVVRDSEIAKETQAKVLEALIKEDSEVKAKEVLEVDVEIAKYOVLEKLV 664
 DB 1498 KITRGLSEVVRDSEIAKETQAKVLEALIKEDSEVKAKEVLEVDVEIAKYOVLEKLV 1557
 OY 665 IHEQITDLSKYAIGPHVAIAKRLAAGKIVRPGTIIISYIVLKGSGKISDRVILLSEYD 724
 DB 1558 IHEQITDLSKYAIGPHVAIAKRLAAGKIVRPGTIIISYIVLKGSGKISDRVILLSEYD 1617
 OY 725 PKRHKYPPDYIENOVPAVRLIEAEGYRKEDLKQSSKQVGLDAMLK 773
 DB 1618 PKRHKYPPDYIENOVPAVRLIEAEGYRKEDLKQSSKQVGLDAMLK 1666

RESULT 8
 A69312
 DNA polymerase B1 (polB) homolog - Archaeoglobus fulgidus
 C:Species: Archaeoglobus fulgidus
 C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 22-Oct-1999
 C:Accession: A69312
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
 Nature 390, 364-370, 1997
 A:Authors: Usterbach, T.; Cotton, M.D.; Spriggs, T.; Artlich, P.; Kaine, B.P.; Sykes, S.
 Smith, H.O.; Woese, C.R.; Venter, J.C.
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo
 A:Reference number: A69250; MUID:98049343
 A:Accession: A69312
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-781 <KLE>
 A:Cross-references: GB:AE001070; GB:AE000782; NID:92689393; PID:AB90741.1; PID:g265013
 C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 32.3%; Score 1301; DB 2; Length 781;
 Best Local Similarity 37.4%; Pred. No. 1.9e-59;
 Matches 305; Conservative 155; Mismatches 266; Indels 90; Gaps 21;

Y 2 IFTDITTKGKPIIRIKKENGFEKLEDPHOPITYALIKDSAIDEIKATGERHKG 61
 DB 8 LIDADYETIGKAVRMLCKDDGIFVAYVNFDPYFVIGVDE--DILKNATSTRRE 64
 OY 62 IIRVAVDAVKKKKRFLGRDVEWKLIFEPDOPALRGKIREHPAVIDIEYEDIFAKRYL 121
 DB 65 VIKIKSEKQKLTGLGEVEGYIYAHHPKRLDYLSQF--DREDIDIFAKRYL 121
 OY 122 IDKGL-----IPMGDE-----ELKMAFDIEFYHGC-DEFG 153
 DB 122 IDKGLACMGAIIRGEGYIRSYIEKVERIPRMEPELKMVLFQCEMLSSGMEPE 181
 OY 154 KGEIIMSYADEEAKRYITKNIDLPYDVYNSREMIKRVQIVYREKDPVLTYYGDN 213
 DB 182 KDPIIVISVTNDDEIIT-----LTGDERKIKISFVKIKISYDPIIYGYODA 230
 OY 214 FDLPIYLRKAEKLGVTLLGDRKHEPKIHRMGDSFAVEIKHFDLPVVRRTINLP 273
 DB 231 FDMPIYLRKAEKRNWIPLDVGRDSN---VYFRKG---RPKITGLNDVLDIAMRISDIK 284
 OY 274 TYTLEAVEAVLGKTK-SKLGAEIRIAIWETESMKRLAOSYMDARATYELGFEPPME 332
 DB 285 IKKLENAEFLGKRIEADIEAKDIYRWSGEK-EKVLNVARQDANTYIAAEILLPMH 343
 OY 333 AELAKLIGOSVMDVSRSGTGNLVEMYLLRVAYERNEIAPNKPDEEYRRRLTYYLGCV 392

DB 344 YELSMIRLPVDVYTRMGKQVMDLLSEAKKIGELIAPNPENAE-----SYGAFV 396
 OY 393 KEPERGLMENTIYDFRCLPSIITYHNVSPDTLERECKN--YVAPRYGKFECKDPG 450
 DB 397 LEPERGLHENVACLDPEASMYPSIMIAFNISPTY--GCRDQCYE-APEVGKFRKSPDG 452
 OY 451 FIPSTIGLITMROEIKKKM-ATIDPIEKKLDYRORAVKHAHSYGYMGYPARWY 509
 DB 453 FFKRLRLRIEKRRLKELKLNLSPESEYKLLDKOOTLAKLTNLSFGYKMNMLARWY 512
 OY 510 KCAESVTAMGRHYIEMTKEIEKFGFVLVADTDFGYATIPGEPETIKKAKKFLKY 569
 DB 513 HPCAEATVAMGRHYIEMTKEIEKFGFVLVADTDFGYATIPGEPETIKKAKKFLKY 569
 OY 570 INSKPLLELEBEFYLRGFPVAKKRYAVIDEBIRITRGLSEVVRDSEIAKETQAKV 629
 DB 568 LHLELP--TQLEVDYYSALFVEKRRYAGLGEDRLVKGLEVRGOWCELAQVOREV 625
 OY 630 LEALIKEDSVKAEIVADVVEEIAKYOVPLEKTIHQITDLSKYAIGPHVAIAKRL 689
 DB 626 IEVLIKENKPKALSLVADVILIRIEKGVSLSEVYIYGTGKPKRYSMOAHVAAALKA 685
 OY 690 AAKGIVRPGTIIISYIVLKGSGKISDR--VILLSEYPPK-----HKYDPDY 735
 DB 686 REMGIYVSSKIGYIVKGSNGIGDRAYPIDLIEDFGENLRITKSGIEIKKIDKDY 745
 OY 736 IENOVPAVRLIEAEGYRKEDLKQSSKQVGLDAMLK 771
 DB 746 IDNITPSVLRILERFGYTEASLK--GSSQMSLDSF 779

RESULT 9
 E64410
 DNA-directed DNA polymerase (EC 2.7.7.7) family B, intein containing precursor - Meth
 N:Contains: DNA endonuclease (EC 3.1.-.-) PI-I; DNA endonuclease (EC 3.1.-.-) PI-II;
 C:Species: Methanococcus jannaschii
 C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Sep-1996
 C:Accession: E64410
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
 .; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
 Science 273, 1058-1073, 1996
 A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
 A:Reference number: A6300; MUID:96337999
 A:Accession: E64410
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1634 <BUL>
 A:Cross-references: GB:U67532; GB:U77117; NID:g1591559; PID:g1591563; TIGR:MJ0885; PI
 C:Genetics:
 A:Map position: REV816304-811400
 A:Start codon: TTG
 C:Function: <DPL>
 A:Description: as DNA-directed DNA polymerase, catalyzes the polymerization of DNA at
 C:Function: <EN1>
 A:Description: as DNA endonuclease PI-MjI, catalyzes the hydrolysis of internal phos
 C:Function: <EN2>
 A:Description: as DNA endonuclease PI-MjII, catalyzes the hydrolysis of internal pho
 C:Superfamily: DNA-directed DNA polymerase KOD
 C:Keywords: endonuclease; hydrolase; nucleotidyltransferase; protein splicing
 F:1-425,795-882,1359-1634/Product: DNA-directed DNA polymerase family B #status predi
 F:1-425,795-882,1359-1634/Product: DNA-directed DNA polymerase family B #status predi
 F:426-794/Product: DNA endonuclease PI-I (pol B intein 1) #status predicted <MjI>
 F:783-882/Product: DNA-directed DNA polymerase family B #status predicted <MjII>
 F:883-1358/Product: DNA endonuclease PI-II (pol B intein 2) #status predicted <MjII>
 F:1558-1634/Product: DNA-directed DNA polymerase family B #status predicted <MjII>
 F:425-795/Cross-link: peptide (Arg-Ser) #status predicted
 F:882-1359/Cross-link: peptide (Asn-Ser) #status predicted

Query Match 32.1%; Score 1291; DB 2; Length 1634;
 Best Local Similarity 22.7%; Pred. No. 1.6e-58;


```

      116 FAKRYLIDKGLIPE-----GDEELKLMADI 142
      117 FRRRLIDKSLVPEELEEFGVEVDASVTTDVRTVEVTRVOSTGSGHGLDLSFDI 176
      143 ETPYHEG-DEFGKEIIMISYA-DEEARVITMKNIDLPHYDVVSNREMIKRFQVIRE 200
      177 EVRPHGMPDEKDEIIVAGVAGNMGVSVISTGMDHDLDFEYVDEDELELREFEYID 236
      201 KDPVLTLYNGDNFDPYLIRAEKGLVTLILGRDKEHPEKIRHMGDSF--AVEIKRI 258
      237 KKPILVGYNSDNDFPYTRRAAILGAEILDIGWDS--KIRMRGRFANATAIKGTV 292
      259 HFDEFPVRRITNPTYLEAVYEAVLGKTSKIGAEIAIWEESMKLAQVSMDA 318
      293 HVDLPVRRKRMNDRITLERNYQELFGEKIDLPGRMLMEYWDRLDELFRYSLDV 352
      319 RAYELGKEFPPEAEALAKLIGQSVMDVSRSTGNLVWYLLRAVERNELAPNKDEEE 378
      353 VATHRIEKLILPMLDELRLVGOPLFDISRMATGQAEWFLVRAKYOGELVPNKPSGD 412
      379 YRRRLRTTYLGGYVKEPERGMENTTYLDFRCLYPSIIVTNVSPDYLERRGCKNYDAP 438
      413 FSSRRGRRAVGYVKEPEKGHENIVQDFRSLPSIISKNISPDITLDEESECYAP 472
      439 IVGKFCDFPGFIPSIILGELITMRKOEIKKMKATIDPIEKKMLDVRORAVKIHANSYG 498
      473 EYGRFRKSPGFPVSYIGELISERVRIKEMKSDDEMEKILWQOEAALKRLANMIG 532
      499 YMGYPKARWYSKECAESVTANGRHVIENTIKEIEKFGKVLVADTDFVATIPG 553
      533 VYGSRFRWYSMECAEALTANGRDYIKTITAE-PEGHIVYADTDFEATYR 586

```

```

RESULT 11
JC7382
DNA-directed DNA polymerase (EC 2.7.7.7) B3 - Sulfurisphaera ohwakuensis
N:Alternate names: DNA polymerase alpha, beta, gamma; DNA polymerase I, II, III
C:Species: Sulfurisphaera ohwakuensis
C>Date: 17-Nov-2000 #sequence_revision 17-Nov-2000 #text_change 31-Dec-2000
C:Accession: J07382
C:Title: Sequence analysis of three family B DNA polymerases from the thermacidophilic
R:Iwai, T.; Kurosawa, N.; Itoh, Y.H.; Kimura, N.; Horiiuchi, T.
DNA Res. 7, 243-251, 2000
A:Title: Sequence analysis of three family B DNA polymerases from the thermacidophilic
C:Reference number: J07380
A:Accession: J07382
A:Molecule type: DNA
A:Residues: 1-781 <IWA>
A:Cross-references: DDBJ:AB032376
C:Comment: This enzyme has both 3'-5' exonuclease and polymerase activities, and plays a
C:Genetics:
A:Gene: B3
C:Keywords: DNA replication; exonuclease; metal binding; nucleotidyltransferase

```

```

Query Match          29.9%; Score 1204.5; DB 2; Length 781;
Best Local Similarity 36.6%; Pred. No. 1.7e-54;
Matches 294; Conservative 141; Mismatches 282; Indels 87; Gaps 21;

```

```

      2 IFDPTDTKDGPIIRIFK-KENGEFKIELDPHFQPIYALALKDSADIEIK-AIKGER 58
      7 ILDESVDVERKPVYIWIWIDKGNRVYLLKEKFRPYALVDSYNDIEIRKEIKLSK 66
      59 HGKIVRVADVAVKKKFLGRVDEWKLIFEHPOVPALRGKIREHRAVIDIYEDIPAK 118
      67 PYSPTSIDV--EKKKTFGSPVKALKIETVPAVVRVYRDEVAKIKGVSVLEADIRYM 124
      119 RYLLDKGLIPE-----EGDE-----ELKLAFDIEFTYH 147
      125 RYSDIDILKPPYWEAEVEEIKENNRKVKYELKIKNLKIEDKIPELKVLAFDIEY-- 182

```

```

      148 EGDEFG-----KGEIIMISYADEEARVITMKNIDLPHYDVVSNREMIKRFQVIREKD 202
      183 --NKYGSNPRRDPYIILIGVMTKEGGKQELAKYD-----DLRAIREFINVQTD 231
      203 PDVLTLYNGDNFDPYLIRAEKGLVTLILGRDKEHPEKIRHMGDSFAVEIKRIHFD 262
      232 PDIIIVGINNEDWPLYLERANIRGIRLDVGR-RVNGEPS--QGYYCHYSTIGRLNVDL 287
      263 FPPVRRITNPTYLE--AVYEAVLGKTSKIGA-ELIAIWEESMKLAQVSMEDAR 319
      268 YGEAOSIOEAVKVTLENADYGLVLPKEREKRTIVEMDIPKYVDEKRDILKLYNLDDAK 347
      320 ATYELGKEFPPEAEALAKLIGQSVMDVSRSTGNLVWYLLRAVERNELAPNKDEEEY 379
      348 SAYLLGEVIFPGIELFTISGLPLDQLMASAGHVEWMLMEAKVYNLLINK-BERY 406
      380 RRRRLTYLGGYVKEPERGMENTTYLDFRCLYPSIIVTNVSPDYLERRGCKNYDAP 439
      407 E-----SYEGGLVISPGLGHEEVVLDFFSSMYPSIMIKYNIGPDLVGECECQWSP- 460
      440 VGYKFCDFPGFIPSIILGELITMRKOEIKKMKATIDPIEKKMLDVRORAVKIHANSYG 499
      461 VGHKTRKEPPLGKLVLEKLIQERKEVKKMKETIDEDKRVLDKQRLAKYMANAFGY 520
      500 MGYPRARWYSKECAESVTANGRHVIENTIKEIEKFGKVLVADTDFVATIPGKPEPI 559
      521 MGMLGARWYSKEGAAVYAMGQIISDSAKIAKER-GFTVIYIGDDISFVKGGD----- 574
      560 KKKAKEPLKYNISKLPGLLELEYEGFYLRGFVA-KKRAYIDEGSRITTRGLEVVRDW 618
      575 --INSLITEISSKFG--LEIKIDIKYRVFETENKRRAGLIEDGKIDIVGEFAVRGDW 629
      619 SEIAKETAKVLEALIKEDSVKAEVIEKDVVEAEIKAOVPEKIVIHQIKOLISEYKA 678
      630 CDIAQOVQTNVIELKSGKVEDAIKVKASYFDLRIRNFRIEDLLIMKTIDKLNDEIDV 689
      679 IGPVYAIKRLAAKGIKVRPGTIISYIVLRGSGISDRY--ILLSEYDPKKHKYDPDY 735
      690 TAPVYAAKRLAAKAGVLYSKVIGYIVYVKGSGISDKAEPFLVKE----KNKIDVEY 745
      736 IENQVLPVAVLRLLEFGYRKEDLK 759
      746 IDKQIIPALRLLEFGVKESSLK 769

```

```

RESULT 12
B56277
DNA-directed DNA polymerase (EC 2.7.7.7) II - Pyrodicticum occultum
C:Species: Pyrodicticum occultum
C>Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 20-Jun-2000
C:Accession: B56277
R:Uemori, T.; Ishino, Y.; Doi, H.; Kato, I.
J. Bacteriol. 177, 2164-2177, 1995
A:Title: The hyperthermophilic archaeon Pyrodicticum occultum has two alpha-like DNA p
A:Reference number: A56277; MUID:95238290
A:Accession: B56277
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-803 <UE>
A:Cross-references: GB:D38574; NID:g807829; PIDN:BA07580.1; PID:g807830
C:Superfamily: herpesvirus DNA-directed DNA polymerase
C:Keywords: nucleotidyltransferase

```

```

Query Match          29.6%; Score 1190; DB 2; Length 803;
Best Local Similarity 33.7%; Pred. No. 9.8e-54;
Matches 281; Conservative 164; Mismatches 278; Indels 110; Gaps 23;

```

```

      1 MIFPDY-----ITKDGKPIIRIFK-KENGEFKIELDPHFQPIYALALKD--D 45
      8 VLLDSYVILKEPVIYIIMGTLIDGKRYVL-----LDHRRPYFALALARGYE 55
      46 SAIDEIKAIKGERHGIYR--VVDVAVKVKKFLGRDVEWKLIFEHPOVPALRGKIREH 103

```

```

Db 56 DWVEELAA--SIRLSVSKSPITDAKPLDKRFRGRKRAVKITTMIPESVRHYREAVAKKI 113
OY 104 PAVIDIYEIDFPAKRYLLDKGL-----IPME-----GDPE-----134
Db 114 EEVEESLEADIFAMRYLLDKRLPTTYRIRIPEVDAGRPRGRVDRVYVAGDPEPLADI 173
OY 135 -----LKLMAFIETFFHEGD-EFGKEIIMISYADE-BAVITMKNIDLPRVDVVS 185
Db 174 RTIDLPRLMVLAFIDIEVYSRSGSPNAPRPVITVSLRDESGERLIEAGCHD-----225
OY 186 NREMIKRRVOYVREKDPVLLTYNGDNDFLLYLLKRAKLGVTLLGDKHEPEPKIR 245
Db 226 -DRRLREVEVYRAFPDPIIVGYSNHPDWYLMERARLGIKLDVTR-RVGAETTSV 283
OY 246 MGDSPFAVEIKGRHFDLFVVRRTINLPYTLAEVAVELGKTSKGAE--IAAIME 302
Db 284 YGH---VSVOGRNLNDLYDAEMPEIKKTELEVAEYICVMKKSERYIEMRIPEYD 340
OY 303 TRESKKLAQYSEMDARATYELGKEFFPMABELAKLIGOSVMDVSRSTGNLVEWYLLRV 362
Db 341 DEKKHQLLERYALDVRATYGLAEKKLPALQISTYGVPLDQVGMVGFRLEWYLLMRA 400
OY 363 ATERNEMLANRDEEYRRRLRTTYLIGYVKEPERGLMNTITYLDRCLYPSIYTHNS 422
Db 401 AYDMNELVPRNRE-----RGESEYKGAVALKPLKGVHENVVLDSSMYPSTIMIKYNG 454
OY 423 PDLT--ERECKKNYD--VAPIVGKFCDFPFISIGELITMROELKKMKKA-TIDP 476
Db 455 PRTIYDDPSECKYGGCYAPVGVHFRSPGPFETVLENLKLRLQVKEKKKEPPPS 514
OY 477 IKKMLDYQORAVKLHANSYGYMGYPRKMYSKCEASVYAMGRHYIEMTKEIEKFG 536
Db 515 PEYRLYDEQOKLKLVLANSYGYMGSHARWYCKRCAEAVTAMGRNLI-LTAIEYARKIG 573
OY 537 FVVLADTDCGFATIPGEKPEITIKKAKPELYINSKLGLELEGEVYLGGEFV-ARK 595
Db 574 LKVIYIGDTSLLVYVD-----KEKVEKLEIEVEKELO--PEIKIDKTYKKVFEAK 624
OY 596 RYAVIDEGRITTRGLEVVRDMSEIAKETQAKVLEALIKEDSVKEAVEIVDVEELAK 655
Db 625 RVGLLEGRIDIVGEAVRGMCELAKEVGEKAEIVLNTGVNDAKAIYIRVILQMLE 684
OY 656 VOYPLEKLYHQITKDLSEYKAIGPVAIAKRLAKGKVRGTIISTIVLGSCKID 715
Db 685 GVPYITKLIMKTLSEYKREYEDAPHYMAARMKKAGYVSGDVGVIYVKSQSVSS 744
OY 716 RYILLSEYDPKKHKYDPOYIENOVLPAYLRILEAGYKEDLKQSSKOVGL 768
Db 745 RAYPYFMDPS--TIDVNYIDHQIVPALRILISTFGYTEKOLKAATVQRSL 795

```

RESULT 13

E72515
 C:probable DNA-directed DNA polymerase APE2098 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: E72515
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamaya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum A:Reference number: A72450; MUID:99310339
 A:Accession: E72515
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-784 <KAW>
 A:Cross-references: DDBJ:AP000063; NID:95105654; PIDN:BAAB1109.1; PID:95105797
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE2098
 C:Superfamily: herpesvirus DNA-directed DNA polymerase

Query Match 28.3%; Score 1138; DB 2; Length 784;
 Best Local Similarity 34.6%; Pred. No. 4,5e-51;
 Matches 282; Conservative 137; Mismatches 287; Indels 110; Gaps 19;

```

OY 11 DCKPIIRIRKKEGKERKIELDHPQYIYALLKDDSAIDEIKAIKGERGKIYRVYDANK 70
Db 16 DGSRRVVFY--GEPR-----PYFVLPDGSVGLQDLAMIRRLSPSSPIISVER 63
OY 71 VKKFLGRVVEWKLTFEHPDVPALRGKIREHPAVIDIYEDVIPAARLYLDKGLPM- 129
Db 64 VRRRFIGREVEALKVTLTPASVREYREAVRRLGVRDVLADIPALRPIIDFNLYPAR 123
OY 130 -----EGD-----EELKMAFIETFFYH-EGDE 152
Db 124 WYVAEVEVAVPHGVSVDRAYTLSDIREDETRIQEDPLKGLVMAFDIEVYSKMRTPDP 183
OY 153 GKGEIIMISYAD-----EEBARIYTKNIDLPRVDVVSNEREMIKRVOYVREKDP 204
Db 184 KDPVIMIGLQAGSEIELEADR-----SDKKYIAGVEVRKSIDP 227
OY 205 VLTITNGDNFDLPYLTKRAKLGVTLLGRDKEPEPKIHRMGDSPAVEIKGRHFDLP 264
Db 228 VIVGYQNPNFDPYLYERARVGVKLAVGRSVEPQGLYG-----HYVSGRLNLDLD 282
OY 265 VVRRTINLPYTLAEVAVLGKTK--SKIGAE--EIAIWEESMKLAQYSEMDARA 320
Db 283 FAELHEVAVKVTLEEVAD-VLGAVKIGERYTLEMMQIGEWMDPSKREILRKLRDVRS 341
OY 321 TYELGKEFFPMELAKLIGOSVMDVSRSTGNLVEWYLLRVAYENNELANRPDEEVR 380
Db 342 TNGLAKFLPFGAELSOVGLPLDQVMAASVGRLLRMLREAKIGELVAPNVERSEGR 401
OY 381 RRLTTYLGGYKPEPERGLMENTITYLDFRCLYPSIYTHNSPDLTERECCKNYD-----435
Db 402 -----YACAIYLRPKRGVHEDIAVLDFASMYNIMVKNVGGDILVRG--EEYSGEEVY 454
OY 436 VAPIVGKFCCKDPGFIPISELGITMROELKKMKKA-TIDPIEKKMLDYQORAVVLHAN 494
Db 455 TAPVGVHKKFRKSPGPFKTLERFLSMRQIRSEMKHPPDSPEYKLLDEROKAIFLKN 514
OY 495 SYVGYGPKRMYSKCEASVYAMGRHYIEMTKEIEKFGKGYUADPGFYATIPGE 554
Db 515 ASYGYMGPHARMYKCECEBAVYAMGRSITIRAKAGF-LGLEVIYIGDTSLFVKNDEP 573
OY 555 KEPTIKKAKPELYINSKLGLELEGEVYLGGEFV-AKKRYAVIDEGRITTRGLEV 613
Db 574 KYERLLRFVEBELGF-----DIKVDKYVRVFEFTAKKRYVGLTVDGKIDVGVGEA 624
OY 614 VRDMSEIAKETQAKVLEALIKEDSVKEAVEIKDYVBEIAKYQVPLEKLYHQITKDL 673
Db 625 VRGDMSELAKETQFKVAEIVLKTGSVDEADVVRNITIEKLRQVDMKRLVILKTLTRP 684
OY 674 SEYKAIGPVAIAKRLAAGIKVRPGTIISTIVLRSKGSIDSVIILLSEDPKKHKYDPP 733
Db 685 SMIEARQPHVTALILMERGIKVEPAKIGIVYTKSSGPIYTRA--KPYFMASKEEDVDE 742
OY 734 YYIENQVLPAVLRIILEAFGRKEDLKQSSKOVGLD 769
Db 743 YYVDKQVYVPAALRIILDFYFVTEKRLKGGGRQSYLLD 778

```

RESULT 14

T05731
 DNA-directed DNA polymerase (EC 2.7.7.7) delta chain - soybean
 C:Species: Glycine max (soybean)
 C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 18-Jun-1999
 C:Accession: T05731
 R:Collins, J.T.B.; Cannon, G.C.; Heinhorst, S.
 submitted to the EMBL Data Library, August 1997
 A:Reference number: 215439
 A:Accession: T05731
 A:Status: preliminary; translated from GB/EMBL/DDBJ

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 28, 2002, 14:17:02 ; Search time 14.68 Seconds
(Without alignments)
1287.834 Million cell updates/sec

Title: US-09-803-165-34

Perfect score: 4026
Sequence: 1 MIFPDYIKDKGPIIRIPK.....KEDLYQSSKQVGLDMLKK 774

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgnt_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgnt_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgnt_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgnt_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgnt_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgnt_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3225.5	80.1	774	3	US-08-902-632-2
2	3222.5	80.0	774	3	US-08-073-354-1
3	3222.5	80.0	774	3	US-08-656-005A-1
4	3222.5	80.0	774	4	US-09-073-259-1
5	3222.5	80.0	774	4	US-09-363-095-1
6	3222.5	80.0	774	4	US-09-418-027-1
7	3209.5	79.7	778	2	US-08-906-925-4
8	3151	78.3	775	1	US-07-966-278-1
9	3151	78.3	775	1	US-08-424-921-1
10	3151	78.3	775	1	US-08-556-355A-1
11	3151	78.3	775	2	US-07-803-627A-1
12	3139	78.0	779	2	US-08-688-649-37
13	3139	78.0	779	1	US-08-375-134-12
14	3139	78.0	779	5	PCT-US95-15263-12
15	2502.5	62.2	1022	1	US-08-271-364A-8
16	2502.5	62.2	1022	2	US-08-222-715B-27
17	2069	51.4	1019	1	US-08-271-364A-7
18	2069	51.4	1019	2	US-08-232-715B-26
19	1595	39.6	396	1	US-08-229-284A-2
20	1403	34.8	788	2	US-08-907-166-6
21	1194	29.7	803	2	US-08-907-166-4
22	1191	29.6	803	1	US-08-062-368-4
23	1180	29.3	803	1	US-08-062-368-2
24	611	15.2	1107	1	US-08-366-577-2
25	591	14.7	877	5	PCT-US96-00005-2
26	588.5	14.6	762	2	US-08-907-166-8
27	588.5	14.6	762	2	US-08-907-166-10

28	585.5	14.5	1462	3	US-07-792-600-31	Sequence 31, Appl
29	585.5	14.5	1462	3	US-09-157-021-31	Sequence 31, Appl
30	585.5	14.5	1462	3	US-09-156-842-31	Sequence 31, Appl
31	479.5	11.9	1009	2	US-08-680-326-31	Sequence 31, Appl
32	467	11.6	783	1	US-08-101-593-6	Sequence 6, Appl1
33	467	11.6	783	1	US-08-465-995A-6	Sequence 6, Appl1
34	467	11.6	783	2	US-08-465-994C-6	Sequence 6, Appl1
35	463.5	11.5	1015	2	US-08-680-326-32	Sequence 32, Appl
36	463	11.5	837	2	US-08-680-326-117	Sequence 117, App
37	461	11.5	1012	2	US-08-680-326-34	Sequence 34, Appl
38	458	11.4	1094	2	US-08-680-326-40	Sequence 40, Appl
39	457	11.4	108	3	US-09-073-354-10	Sequence 10, Appl
40	457	11.4	108	3	US-08-656-005A-10	Sequence 10, Appl
41	457	11.4	108	4	US-09-073-259-10	Sequence 10, Appl
42	457	11.4	108	4	US-09-363-095-10	Sequence 10, Appl
43	457	11.4	108	4	US-09-418-027-10	Sequence 10, Appl
44	433.5	10.8	1008	2	US-08-680-326-30	Sequence 30, Appl
45	426.5	10.6	1097	2	US-08-680-326-39	Sequence 39, Appl

ALIGNMENTS

RESULT 1
US-08-902-632-2
Sequence 2, Application US/08902632
Patent No. 6008025
GENERAL INFORMATION:
APPLICANT: KOMATSUBARA, Hideyuki
APPLICANT: KITABAYASHI, Masao
APPLICANT: KAWIMURA, Hideki
APPLICANT: KAWAKAMI, Bunsei
APPLICANT: KAWAMURA, Yoshihisa
APPLICANT: TAKAGI, Masahiro
APPLICANT: IMANAKA, Tadayuki
TITLE OF INVENTION: Modified Thermostable DNA Polymerase,
TITLE OF INVENTION: and DNA Polymerase Composition for Nucleic Acid
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1 Broadway
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10004
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.50 inch floppy disk, 1.44 MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/902.632
FILING DATE: Concurrent Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 198911/96
FILING DATE: 29-JUL-1996
APPLICATION NUMBER: JP 200446/96
APPLICATION NUMBER: 30-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: GREASON, Edward W.
REGISTRATION NUMBER: 18,918
REFERENCE/DOCKET NUMBER: 2418/7
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-425-7200
TELEFAX: 212-425-5288
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein


```

QY 421 VSPDTEREGCKNDVAVIVGKCKDPGFPISILGELITMROEIKKKMATIDPIEKK 480
Db 418 VSPDTEREGCKNDVAVIVGKCKDPGFPISILGELITMROEIKKKMATIDPIEKK 477
QY 481 MLDYRQRAVKIHAANSYGYMGYPRKAWYKSCAASVTAMGNHYIEMTKEIEEKGFEVYL 540
Db 478 LLDYRQRAIKILANSYGYMGYPRKAWYKSCAASVTAMGNHYIEMTKEIEEKGFEVYL 537
QY 541 YADIDGFATIPGEPFATIPGADAEIVKKAKEFLYINSKLPGLLEIEYEGFYLGFPVAKKRYAVI 600
Db 538 YSDIDGFATIPGADAEIVKKAKEFLYINSKLPGLLEIEYEGFYLGFPVAKKRYAVI 597
QY 601 DEGRITTRGLEEVVRRDMSEIAKETQAKVLEAILKEDSEKAVELVKDVEIAKYYOVL 660
Db 598 DEGRITTRGLEEVVRRDMSEIAKETQAKVLEAILKEDSEKAVELVKDVEIAKYYOVL 657
QY 661 EKLVIHEQITDLSYKAIKGVHVAIAKRLAKGKIKVRGTTIISYIVLKGSKISRVLL 720
Db 658 EKLVIHEQITDLSYKAIKGVHVAIAKRLAKGKIKVRGTTIISYIVLKGSKISRVLL 717
QY 721 SEYDPKHKHYDPDYIENQVPAVLRILIEAFGRKEDLKYSKQVGLDAMLK 773
Db 718 DEFPDTHKHYAEYIENQVPAVLRILIEAFGRKEDLKYSKQVGLDAMLK 770

RESULT 3
US-08-656-005A-1
: Sequence 1, Application US/08656005A
: Patent No. 6054301
: GENERAL INFORMATION:
: APPLICANT: KITABAYASHI, Masao
: APPLICANT: ARAKAWA, Taku
: APPLICANT: INOUE, Hiroaki
: APPLICANT: KAWAKAMI, Bunsei
: APPLICANT: KAWAMURA, Yoshihisa
: APPLICANT: IMANAKA, Tadayuki
: APPLICANT: TAKAGI, Masahiro
: TITLE OF INVENTION: A Method of Amplifying Nucleic
: TITLE OF INVENTION: Acid and A Reagent Therefor
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenyon & Kenyon
: STREET: 1025 Connecticut Avenue, N.W., Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WordPerfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/656, 005A
: FILING DATE: 24 MAY 1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: JP 134096/95
: FILING DATE: 31 MAY 1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Toffenetti, Judith L.
: REGISTRATION NUMBER: 39,048
: REFERENCE/DOCKET NUMBER: 2418/3
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-429-1776
: TELEFAX: 202-429-0796
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 774 amino acids
: TYPE: amino acid
: STRANDEDNESS: double
: TOPOLOGY: linear

```

```

: MOLECULE TYPE: protein
: US-08-656-005A-1
Query Match 80.0%; Score 3222.5; DB 3; Length 774;
Best Local Similarity 77.4%; Pred. No. 6,56-243;
Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIPDIDYTKGKPIRIRFEKKENGFEKTELDPHOPYIYALLKDSADIEKAIKGEHNG 60
Db 1 MIPDIDYTKGKPIRIRFEKKENGFEKTELDPHOPYIYALLKDSADIEKAIKGEHNG 60
QY 61 KIVRVYDAVKKKKRFLGHDVEMWKLIFEPHODVPALRGIKREHPAVIDIYEYDIPFARY 120
Db 61 KIVRVYDAVKKKKRFLGHDVEMWKLIFEPHODVPALRGIKREHPAVIDIYEYDIPFARY 120
QY 121 LIDKGLIPMEDEBELKMAFDIETPFYHEGDEKGEIIMISYADEEARKVITWKIDLPY 180
Db 121 LIDKGLIPMEDEBELKMAFDIETPFYHEGDEKGEIIMISYADEEARKVITWKIDLPY 180
QY 181 VDVVSNREEMIKRFVQIYREKDPDVLITVNGDNFPLILIRAEKLGVTLLGRDKHEPE 240
Db 181 VDVVSNREEMIKRFVQIYREKDPDVLITVNGDNFPLILIRAEKLGVTLLGRDKHEPE 240
QY 241 PKIHRMGDSFAVEIKGRIFPDLFPVVRTINLPYTTLEAVEAVLGTGKSKGAEIYAI 300
Db 241 PKIHRMGDSFAVEIKGRIFPDLFPVVRTINLPYTTLEAVEAVLGTGKSKGAEIYAI 300
QY 301 KPIRGMGRFAVEYKGRHFDLPYIRRTINLPYTTLEAVEAVLGTGKSKGAEIYAI 360
Db 301 KPIRGMGRFAVEYKGRHFDLPYIRRTINLPYTTLEAVEAVLGTGKSKGAEIYAI 360
QY 361 RFAVERNEELNLRVARYSMEDAKVYELGKEFLPMEAOQLSRIGOSLMDVSRSSGTNLEWELL 358
Db 361 RFAVERNEELNLRVARYSMEDAKVYELGKEFLPMEAOQLSRIGOSLMDVSRSSGTNLEWELL 358
QY 359 RFAVERNEELNLRVARYSMEDAKVYELGKEFLPMEAOQLSRIGOSLMDVSRSSGTNLEWELL 358
Db 359 RFAVERNEELNLRVARYSMEDAKVYELGKEFLPMEAOQLSRIGOSLMDVSRSSGTNLEWELL 358
QY 421 VSPDTEREGCKNDVAVIVGKCKDPGFPISILGELITMROEIKKKMATIDPIEKK 480
Db 418 VSPDTEREGCKNDVAVIVGKCKDPGFPISILGELITMROEIKKKMATIDPIEKK 477
QY 481 MLDYRQRAVKIHAANSYGYMGYPRKAWYKSCAASVTAMGNHYIEMTKEIEEKGFEVYL 540
Db 478 LLDYRQRAIKILANSYGYMGYPRKAWYKSCAASVTAMGNHYIEMTKEIEEKGFEVYL 537
QY 541 YADIDGFATIPGEPFATIPGADAEIVKKAKEFLYINSKLPGLLEIEYEGFYLGFPVAKKRYAVI 600
Db 538 YSDIDGFATIPGADAEIVKKAKEFLYINSKLPGLLEIEYEGFYLGFPVAKKRYAVI 597
QY 601 DEGRITTRGLEEVVRRDMSEIAKETQAKVLEAILKEDSEKAVELVKDVEIAKYYOVL 660
Db 598 DEGRITTRGLEEVVRRDMSEIAKETQAKVLEAILKEDSEKAVELVKDVEIAKYYOVL 657
QY 661 EKLVIHEQITDLSYKAIKGVHVAIAKRLAKGKIKVRGTTIISYIVLKGSKISRVLL 720
Db 658 EKLVIHEQITDLSYKAIKGVHVAIAKRLAKGKIKVRGTTIISYIVLKGSKISRVLL 717
QY 721 SEYDPKHKHYDPDYIENQVPAVLRILIEAFGRKEDLKYSKQVGLDAMLK 773
Db 718 DEFPDTHKHYAEYIENQVPAVLRILIEAFGRKEDLKYSKQVGLDAMLK 770

RESULT 4
US-09-073-259-1
: Sequence 1, Application US/09073259
: Patent No. 6143536
: GENERAL INFORMATION:
: APPLICANT: IMANAKA, Tadayuki
: APPLICANT: TAKAGI, Masahiro
: APPLICANT: MORIKAWA, Masaaki
: TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Kenyon & Kenyon

```

STREET: 1025 Connecticut Avenue, N.W., Suite 600
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5" Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Wordperfect 6.1 Windows
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/073,259
 FILING DATE: Concurrent Herewith
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/656,005
 FILING DATE: 24 MAY 1996
 APPLICATION NUMBER: JP 134096/95
 FILING DATE: 31 MAY 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Toifenetti, Judith L.
 REGISTRATION NUMBER: 39,048
 REFERENCE/DOCKET NUMBER: 2418/10
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-428-1776
 TELEFAX: 202-429-0796
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 774 amino acids
 TYPE: amino acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-073-259-1

Query Match 80.0%; Score 3222.5; DB 4; Length 774;

Best Local Similarity 77.4%; Pred. No. 6.5e-243;

Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDDYIRKDKPIIRIRKKEGKFKIELDPHFQPIYALAKDSADIEIKAGERHG 60
 DB 1 MLDDYITRDEGKPIYIRIRKKEGKFKIELDPHFQPIYALAKDSADIEIKAGERHG 60
 QY 61 KIVRVVDAVKKKELGRVDEYWKLFEPHODVPALRGKIREHPAVIDIYEDIPAKRY 120
 DB 61 TVVTVKRVKVKOKKELGRVDEYWKLFEPHODVPALRGKIREHPAVIDIYEDIPAKRY 120
 QY 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKELIMISYADEEARAVITWKNDLPY 180
 DB 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKELIMISYADEEARAVITWKNDLPY 180
 QY 181 VDVVSNERMIKRFVQIVREKDPDLITYNGDNFDPYLIRKAEKLGVTLLGRDKEHP 240
 DB 181 VDVVSNERMIKRFVQIVREKDPDLITYNGDNFDPYLIRKAEKLGVTLLGRDKEHP 240
 QY 241 PRIHRGDSFAVEIKRHFDELFPVVRTINLPYTLLEAVYAVLGKTSKGAELIAI 300
 DB 241 PRIHRGDSFAVEIKRHFDELFPVVRTINLPYTLLEAVYAVLGKTSKGAELIAI 300
 QY 239 PRIORGRDFAVEVGRIFHFDLYPIVRTINLPYTLLEAVYAVLGKTSKGAELIAI 298
 DB 239 PRIORGRDFAVEVGRIFHFDLYPIVRTINLPYTLLEAVYAVLGKTSKGAELIAI 298
 QY 301 WTEESMKKLAOYSMEDARATYELGKEFFPMAELAKLIGOSVMPVSSSTGNLVEMVLL 360
 DB 301 WTEESMKKLAOYSMEDARATYELGKEFFPMAELAKLIGOSVMPVSSSTGNLVEMVLL 360
 QY 361 RAYERNELAPKPKDEEYRRRLRTTYLGKYKPEPGRGMENTIYLDFFCLPSIIVTN 420
 DB 361 RAYERNELAPKPKDEEYRRRLRTTYLGKYKPEPGRGMENTIYLDFFCLPSIIVTN 420
 QY 421 VSPDPLERGCCKNYDAPVIGYKFCNDPFGFIPISILGELITROEIKKKMAKATIDPIEK 480
 DB 421 VSPDPLERGCCKNYDAPVIGYKFCNDPFGFIPISILGELITROEIKKKMAKATIDPIEK 480
 QY 418 VSPDPLERGCCKNYDAPVIGYKFCNDPFGFIPISILGELITROEIKKKMAKATIDPIEK 477
 DB 418 VSPDPLERGCCKNYDAPVIGYKFCNDPFGFIPISILGELITROEIKKKMAKATIDPIEK 477
 QY 481 MLDYKORAVKLANSTYGYMGYPKARWYSKCAESVTANGRHYIEKTIKEIEKGEKVL 540
 DB 481 MLDYKORAVKLANSTYGYMGYPKARWYSKCAESVTANGRHYIEKTIKEIEKGEKVL 540

DB 478 LLDYKORAVKLANSTYGYMGYPKARWYSKCAESVTANGRHYIEKTIKEIEKGEKVL 537
 QY 541 VADTGFEVATIPGKPEPTTKRKAELKYNKGLLEGEYGRNGEFAKRYANI 600
 DB 538 YSDTGFEVATIPGKPEPTTKRKAELKYNKGLLEGEYGRNGEFAKRYANI 597
 QY 601 DEGRITTTGLEIVRDMSEIAKETQAKVLAELKEDSVEKAVEIKDVEEIAKYOVL 660
 DB 598 DEGRITTTGLEIVRDMSEIAKETQAKVLAELKEDSVEKAVEIKDVEEIAKYOVL 657
 QY 661 EKLVIHQITKDLSEKKAIGPVHAIKLAAGIVRPTIISYLVKSGISDRVILL 720
 DB 658 EKLVIHQITKDLSEKKAIGPVHAIKLAAGIVRPTIISYLVKSGISDRVILL 717
 QY 721 SEYDPRKHKYDDPYIENOVLPVLRILEAFGRKEDLKQSSQGDAMWLK 773
 DB 718 DEFDPKHKYDDPYIENOVLPVLRILEAFGRKEDLKQSSQGDAMWLK 770

US-09-363-095-1
 Sequence 1, Application US/09363095

Patent No. 6187573

GENERAL INFORMATION:

APPLICANT: IMANAKA, Tadayuki

APPLICANT: TAKAGI, Masahiro

APPLICANT: MORIKAWA, Masaki

TITLE OF INVENTION: DNA Encoding A Thermostable DNA Polymerase

FILE REFERENCE: 2418/11

CURRENT APPLICATION NUMBER: US/09/363,095

CURRENT FILING DATE: 1999-07-30

EARLIER APPLICATION NUMBER: US 09/073,259

EARLIER FILING DATE: 1998-05-06

EARLIER APPLICATION NUMBER: US 08/656,005

EARLIER FILING DATE: 1996-05-24

EARLIER APPLICATION NUMBER: JP 134096/95

EARLIER FILING DATE: 1995-05-31

NUMBER OF SEQ ID NOS: 16

SOFTWARE: WordPerfect 6.1 Windows

SEQ ID NO 1

LENGTH: 774

TYPE: PRT

ORGANISM: Hyperthermophilic archaeon

US-09-363-095-1

Query Match 80.0%; Score 3222.5; DB 4; Length 774;

Best Local Similarity 77.4%; Pred. No. 6.5e-243;

Matches 598; Conservative 83; Mismatches 89; Indels 3; Gaps 2;

QY 1 MIFDDYIRKDKPIIRIRKKEGKFKIELDPHFQPIYALAKDSADIEIKAGERHG 60
 DB 1 MLDDYITRDEGKPIYIRIRKKEGKFKIELDPHFQPIYALAKDSADIEIKAGERHG 60
 QY 61 KIVRVVDAVKKKELGRVDEYWKLFEPHODVPALRGKIREHPAVIDIYEDIPAKRY 120
 DB 61 TVVTVKRVKVKOKKELGRVDEYWKLFEPHODVPALRGKIREHPAVIDIYEDIPAKRY 120
 QY 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKELIMISYADEEARAVITWKNDLPY 180
 DB 121 LIDKGLIPMEGDEELKMAFDIETFYHGEDEFGKELIMISYADEEARAVITWKNDLPY 180
 QY 181 VDVVSNERMIKRFVQIVREKDPDLITYNGDNFDPYLIRKAEKLGVTLLGRDKEHP 240
 DB 181 VDVVSNERMIKRFVQIVREKDPDLITYNGDNFDPYLIRKAEKLGVTLLGRDKEHP 240
 QY 241 PRIHRGDSFAVEIKRHFDELFPVVRTINLPYTLLEAVYAVLGKTSKGAELIAI 300
 DB 241 PRIHRGDSFAVEIKRHFDELFPVVRTINLPYTLLEAVYAVLGKTSKGAELIAI 300
 QY 239 PRIORGRDFAVEVGRIFHFDLYPIVRTINLPYTLLEAVYAVLGKTSKGAELIAI 298
 DB 239 PRIORGRDFAVEVGRIFHFDLYPIVRTINLPYTLLEAVYAVLGKTSKGAELIAI 298
 QY 301 WTEESMKKLAOYSMEDARATYELGKEFFPMAELAKLIGOSVMPVSSSTGNLVEMVLL 360
 DB 301 WTEESMKKLAOYSMEDARATYELGKEFFPMAELAKLIGOSVMPVSSSTGNLVEMVLL 360

```

Db 299 WETGENLERVARYSMEDAKVTEELGKEFLPMEAOLSRLIGOSLMDVSRSSSTGNLWEMLL 358
OY 361 RVAERNEELANRKPDEEYRRRLRTTYIGVYKPERGEMNTINYLDPRCLYPSIIYTHN 420
Db 359 RKAERNEELANRKPDEKELARR-RQSYEGGYVKEPERGLMWNIIYLDPRSLYPSIIYTHN 417
OY 421 VSPDTLREGECKNDVAPIVGYKFCDFPGFIPSLIGELITMROEIKKKMATIDPIERK 480
Db 418 VSPDTLREGECKEYDVAPOVGHRCCKDPGFIPLSLGDLLEEROKIKKKMATIDPIERK 477
OY 481 MLDYRORAVKLHANSYGYMGYPKARWYKSCAASVTAMGRHYIEMTIKEIEKFGKVL 540
Db 478 LLDYRORAIKILANSYGYGYARARWYCKCAESVTAMGREYITMTIKEIEKGYGKVI 537
OY 541 YADTDGFATIPGKPEIKKAKPEIKKYINSKLPGLLEIEYEGFYLRGFFVAKKRAVI 600
Db 538 YSDTDGFEATIPGADAEVTKKAMEFLYINAKLPGLALEIEYEGFYLRGFFVTKKRAVI 597
OY 601 DEGRITTRGLEVVRDMSSEIAKETQAVLEAILKEDSEVERAVEIVKDVVEIAKQVPL 660
Db 598 DEGRITTRGLEIYVRDMSSEIAKETQAVLEAILKDGVEEKAVALIVKEVTEKLSKEYEVP 657
OY 661 EKLVIHQITDLSYKAIQPHVAIAKRLAKGIVRPGTIISYIVLRGSKISDRVILL 720
Db 658 EKLVIHQITDLDKYATGPHVAIVAKRLARGVYKIRPGYISYIVLRGSGRIGDRAIPF 717
OY 721 SEYDPKKHKYDPDYIENQVLPVLRILAEAGYRKREDKYGSSQVGLDAMLK 773
Db 718 DEFDPTRKHKYDAEYIENQVLPVLRILRAGYRKREDLRYOKTROVGLSAMLK 770

```

RESULT 6

US-09-418-027-1

Sequence 1, Application US/09418027

Patent No. 6225065

GENERAL INFORMATION:

APPLICANT: KITABAYASHI, Masao

APPLICANT: ARAKAWA, Taku

APPLICANT: INOUE, Hiroaki

APPLICANT: KAWAKAMI, Bunsei

APPLICANT: KAWAMURA, Yoshihisa

APPLICANT: IMANAKA, Tadayuki

APPLICANT: TAKAGI, Masahiro

APPLICANT: MORIKAWA, Masaaki

TITLE OF INVENTION: A Thermostable DNA Polymerase and Kits for

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kenyon & Kenyon

STREET: 1025 Connecticut Avenue, N.W., Suite 600

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20036

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WordPerfect 6.1 Windows

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/418, 027

CLASSIFICATION:

FILING DATE:

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: 09/073, 354

FILING DATE:

APPLICATION NUMBER: JP 134096/95

FILING DATE: 31 MAY 1995

ATTORNEY/AGENT INFORMATION:

NAME: Toffenetti, Judith L.

REGISTRATION NUMBER: 39, 048

REFERENCE/DOCKET NUMBER: 2418/9

TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 774 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-418-027-1

```

Query Match

80.0%; Score 3222.5; DB 4; Length 774;

Best Local Similarity 77.4%; Pred. No. 6,5e-243;

Matches 598; Conservativity 83; Mismatches 89; Indels 3; Gaps 2;

```

OY 1 MIFDTYITKDKPIIRIFKKEGFEKIEDLPHROPYIYALLKDDSAIDEIKAIKGRHG 60
Db 1 MILDYITTEDGKPVIRIFKKEGFEKIEYDRTEPEFYALLKDDSAIEEVKITAERHG 60
OY 61 KIYVNVDAVKKKKKFLGRDVEVMKLIFFENPODVALNGKIREHRAVIDIYEXDIPAKRY 120
Db 61 TVVTYKRVKQKKFLGRPVEVMKLYFTHPDVPAIRKIREHGAVIDIYEYDIPAKRY 120
OY 121 LIDGLIPMEGDEELKIMAFDIETFYHEGDEFGKEIIMISYADEBARVITMKNDLPY 180
Db 121 LIDGLIPMEGDEELKIMAFDIOTLYHEGDEFEAGPILMISYADEBARVITMKNDLPY 180
OY 181 VDVSNSEREMIKRVOIVREKDPVLLTYNGDNFDLPYLIRAKLGVTLLGDKHEPE 240
Db 181 VDVSSTEREMIKRVLVVKKEKDPVLLTYNGDNFDLPYLRKCKELGINFALGRDGS--E 238
OY 241 PKIRMGDSFAVEIKGRHFDLPFVVRRTINLPYTYLEAVYEAVLGYKTSKIGAEITAAI 300
Db 239 PKIQMGDRFAVEYKGRHFDLPYVIRRTINLPYTYLEAVYEAVFGQPKVEYVABEITPA 298
OY 301 WETEESMKKLAQYSEMDARATYELGKEFFPMEAEIAKLIGOSVMDVSRSSTGNLWEMYL 360
Db 299 WETGENLERVARYSMEDAKVTEELGKEFLPMEAOLSRLIGOSLMDVSRSSSTGNLWEMLL 358
OY 361 RVAERNEELANRKPDEEYRRRLRTTYIGVYKPERGLMWNIIYLDPRCLYPSIIYTHN 420
Db 359 RKAERNEELANRKPDEKELARR-RQSYEGGYVKEPERGLMWNIIYLDPRSLYPSIIYTHN 417
OY 421 VSPDTLREGECKNDVAPIVGYKFCDFPGFIPSLIGELITMROEIKKKMATIDPIERK 480
Db 418 VSPDTLREGECKEYDVAPOVGHRCCKDPGFIPLSLGDLLEEROKIKKKMATIDPIERK 477
OY 481 MLDYRORAVKLHANSYGYMGYPKARWYKSCAASVTAMGRHYIEMTIKEIEKFGKVL 540
Db 478 LLDYRORAIKILANSYGYGYARARWYCKCAESVTAMGREYITMTIKEIEKGYGKVI 537
OY 541 YADTDGFATIPGKPEIKKAKPEIKKYINSKLPGLLEIEYEGFYLRGFFVAKKRAVI 600
Db 538 YSDTDGFEATIPGADAEVTKKAMEFLYINAKLPGLALEIEYEGFYLRGFFVTKKRAVI 597
OY 601 DEGRITTRGLEVVRDMSSEIAKETQAVLEAILKEDSEVERAVEIVKDVVEIAKQVPL 660
Db 598 DEGRITTRGLEIYVRDMSSEIAKETQAVLEAILKDGVEEKAVALIVKEVTEKLSKEYEVP 657
OY 661 EKLVIHQITDLSYKAIQPHVAIAKRLAKGIVRPGTIISYIVLRGSKISDRVILL 720
Db 658 EKLVIHQITDLDKYATGPHVAIVAKRLARGVYKIRPGYISYIVLRGSGRIGDRAIPF 717
OY 721 SEYDPKKHKYDPDYIENQVLPVLRILAEAGYRKREDKYGSSQVGLDAMLK 773
Db 718 DEFDPTRKHKYDAEYIENQVLPVLRILRAGYRKREDLRYOKTROVGLSAMLK 770

```

RESULT 7

US-08-906-925-4

Sequence 4, Application US/08906925

Patent No. 5882904

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

```

Oy 1 MIFPDYITKGRKPIIRFKKENGFKELDPHPQPIYALLKQDSADLKAIGKGEHG 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MILDVDITTEGKRVIRLFKKENGKFKIEHDRTPRPIYALLRQDSKIEEKKITGERHG 60
Oy 61 KIVRVDAVAVKVKKFLGDRVEMWKLIFEPHPODVPALRCKIREHPAVIDIYEYDIPFAKRY 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KIVRVDAVAVKVKKFLGDRVEMWKLIFEPHPODVPALRCKIREHPAVIDIYEYDIPFAKRY 120
Oy 121 LIDKGLIPMEGDEBLKMAFDIETFFYHGDGFGKEIIMISYADEBEARVITWKNIDLPY 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIDKGLIPMEGDEBLKMAFDIETFFYHGDGFGKEIIMISYADEBEARVITWKNIDLPY 180
Oy 181 VDVVSNREMIKRRVOIYREKDPVLTYYNGDNFDLPYLIRAKRLGVTLLGDKHEPE 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VDVVSNREMIKRRVOIYREKDPVLTYYNGDNFDLPYLIRAKRLGVTLLGDKHEPE 240
Oy 241 PKIRHMGDSFAVEIKRIFHFDLPVVRRTINLPYTLLEAVYEAVALGKTSKGAEEIAAI 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PKIRHMGDSFAVEIKRIFHFDLPVVRRTINLPYTLLEAVYEAVALGKTSKGAEEIAAI 300
Oy 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVWDVSRSSGTGNLVWYLL 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVWDVSRSSGTGNLVWYLL 360
Oy 361 RVAERNEMLAPNKPDEEYRRLRTTYLGYYVKKPERGEMINITYLDPRCLYPSIIYTHN 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RVAERNEMLAPNKPDEEYRRLRTTYLGYYVKKPERGEMINITYLDPRCLYPSIIYTHN 420
Oy 421 VSPDTLRECGKNYDAPIYVYKFCDFPGFIPSIILGELITMROEIKKMKATIDPIEKK 480
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VSPDTLRECGKNYDAPIYVYKFCDFPGFIPSIILGELITMROEIKKMKATIDPIEKK 480
Oy 481 MLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 MLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 540
Oy 479 LLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 538
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 479 LLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 538
Oy 541 YADYGFATIPGEPETIKKAKKELFYINSKLPGLLEEGFYLLGPFVAKKRYAVI 600
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 541 YADYGFATIPGEPETIKKAKKELFYINSKLPGLLEEGFYLLGPFVAKKRYAVI 600
Oy 539 YIDYDGLATIPGEGSEBIRKKALEFVYINSKLPGLLEEGFYLLGPFVAKKRYAVI 598
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 539 YIDYDGLATIPGEGSEBIRKKALEFVYINSKLPGLLEEGFYLLGPFVAKKRYAVI 598
Oy 601 DEGRITRGLLEVVRWOMSEIAKETQAKVLEAILKEDSEKAVEIYKDVVEIAKYOYVL 660
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 601 DEGRITRGLLEVVRWOMSEIAKETQAKVLEAILKEDSEKAVEIYKDVVEIAKYOYVL 660
Oy 659 EKLAIYEDITRPLHETKAIQPHVAVAKKLAAGVAKIRGAVIYIVLNGDGPISNRALIA 718
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 659 EKLAIYEDITRPLHETKAIQPHVAVAKKLAAGVAKIRGAVIYIVLNGDGPISNRALIA 718
Oy 721 SEYDPKKHKYDPDYIENQVPAVLRIEAGYRKREDKYOSSKOVGLDAML 772
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 721 SEYDPKKHKYDPDYIENQVPAVLRIEAGYRKREDKYOSSKOVGLDAML 772
Oy 719 EYDPKKHKYDAEYIENQVPAVLRIEAGYRKREDKYOSSKOVGLDAML 770
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 719 EYDPKKHKYDAEYIENQVPAVLRIEAGYRKREDKYOSSKOVGLDAML 770

```

RESULT 9
US-08-424-921-1
Sequence 1, Application US/08424921
Patent No. 554552

GENERAL INFORMATION:
APPLICANT: Machur, Eric A
TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCUS
TITLE OF INVENTION: PURIFIED THERMOSTABLE PYROCOCUS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bingham & Fitting
STREET: 12526 High Bluff Road, Suite 300
CITY: San Diego
STATE: CA
COUNTRY: USA
ZIP: 92130
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,921
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/803,627
FILING DATE: 02-DEC-1991
APPLICATION NUMBER: US 07/620,568
FILING DATE: 03-DEC-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/657,073
FILING DATE: 19-FEB-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/776,552
FILING DATE: 15-OCT-1991
ATTORNEY/AGENT INFORMATION:
NAME: Fitting, Thomas
REGISTRATION NUMBER: 34,163
REFERENCE/DOCKET NUMBER: STG0100P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-792-3680
TELEFAX: 619-792-8477
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 775 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
US-08-424-921-1

```

Query Match 78.3%; Score 3151; DB 1; Length 775;
Best Local Similarity 74.6%; Pred. No. 2.4e-237;
Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

```

Oy 1 MIFPDYITKGRKPIIRFKKENGFKELDPHPQPIYALLKQDSADLKAIGKGEHG 60
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MILDVDITTEGKRVIRLFKKENGKFKIEHDRTPRPIYALLRQDSKIEEKKITGERHG 60
Oy 61 KIVRVDAVAVKVKKFLGDRVEMWKLIFEPHPODVPALRCKIREHPAVIDIYEYDIPFAKRY 120
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 KIVRVDAVAVKVKKFLGDRVEMWKLIFEPHPODVPALRCKIREHPAVIDIYEYDIPFAKRY 120
Oy 121 LIDKGLIPMEGDEBLKMAFDIETFFYHGDGFGKEIIMISYADEBEARVITWKNIDLPY 180
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 LIDKGLIPMEGDEBLKMAFDIETFFYHGDGFGKEIIMISYADEBEARVITWKNIDLPY 180
Oy 181 VDVVSNREMIKRRVOIYREKDPVLTYYNGDNFDLPYLIRAKRLGVTLLGDKHEPE 240
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 VDVVSNREMIKRRVOIYREKDPVLTYYNGDNFDLPYLIRAKRLGVTLLGDKHEPE 240
Oy 241 PKIRHMGDSFAVEIKRIFHFDLPVVRRTINLPYTLLEAVYEAVALGKTSKGAEEIAAI 300
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 PKIRHMGDSFAVEIKRIFHFDLPVVRRTINLPYTLLEAVYEAVALGKTSKGAEEIAAI 300
Oy 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVWDVSRSSGTGNLVWYLL 360
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 WETESMKKLAQYSMEDARATYELGKEFFPMEAEALAKIGOSVWDVSRSSGTGNLVWYLL 360
Oy 361 RVAERNEMLAPNKPDEEYRRLRTTYLGYYVKKPERGEMINITYLDPRCLYPSIIYTHN 420
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 361 RVAERNEMLAPNKPDEEYRRLRTTYLGYYVKKPERGEMINITYLDPRCLYPSIIYTHN 420
Oy 421 VSPDTLRECGKNYDAPIYVYKFCDFPGFIPSIILGELITMROEIKKMKATIDPIEKK 480
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 421 VSPDTLRECGKNYDAPIYVYKFCDFPGFIPSIILGELITMROEIKKMKATIDPIEKK 480
Oy 481 MLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 540
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 481 MLDYRORAVKLANHNSYGYMGYPKARWYSKECAESVTAMGRHYIEMTIKEIEKFGFKVL 540

```

```

DB 479 LLDYRQKAKRLANSFYGYGAKAMWCKECSASTANGRKYIELVMWELEKEFGKVL 538
QY 541 YADTDGFYATIPGEKPEITKKKAKKFLKYINSKLPGLLEVEGEFVLRGFPFAKKRYAVI 600
DB 539 YIDTDGLVATIPGSESEELKKALEFVKYINSKLPGLLEVEGEFVLRGFPFAKKRYAVI 598
QY 601 DEGRITTRGLEVVRDSEIAKETQAKVLEAILKEDSEKAVEIVKQVVEIAKYQVPL 660
DB 599 DEEGVITRGLTEIVRDMSEIAKETQAVLEITLKHGVEEAVRIYKIOKLANYEIRP 658
QY 661 EKLVIHEOITKDSYKAGPHVAIAKRLAAGIKRPGTIISYILRSGKISDPAVL 720
DB 659 EKLAIVEOTRPLHEKKAIPHAAKLAAGVKIKPGMIGVILRGGDPLISRAILA 718
QY 721 SEYDPKHHKYPDDYIENOVLPVLRILEAFGRKEDLKQSSQOYGLDAML 772
DB 719 EYDPRKKHKYDAEYIENOVLPVLRILEGFGYRKEDLRYQKTRQVGLTSM 770

```

RESULT 10

US-08-556-355A-1

Sequence 1, Application US/08556355A

Patent No. 5866395

GENERAL INFORMATION:

APPLICANT: MATHUR, Eric J.

TITLE OF INVENTION: Purified Thermostable Pyrococcus

TITLE OF INVENTION: furiosus DNA Polymerase I

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

ADDRESS: 1300 I Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/556,355A

FILING DATE: 13-NOV-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/424,921

FILING DATE: 19-APR-1995

- PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/803,627

FILING DATE: 02-DEC-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/779,846

FILING DATE: 21-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,073

FILING DATE: 19-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/620,568

FILING DATE: 03-DEC-1990

ATTORNEY/AGENT INFORMATION:

NAME: Barker, M. Paul

REGISTRATION NUMBER: 32,013

REFERENCE/DOCKET NUMBER: 04121,0004-02

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)408-4000

TELEFAX: (202)408-4400

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 775 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-556-355A-1

Query Match 78.3%; Score 3151; DB 2; Length 775;

Best Local Similarity 74.6%; Pred. No. 2.4e-237;

Matches 576; Conservative 100; Mismatches 94; Indels 2; Gaps 1;

```

QY 1 MFEDDYITKCKGPIIRIFKENGGEFKIEDPHFQPIYALAKDSADIEKAIGERHG 60
DB 1 MLDDVYITEEKPIYRIFKENGKFKIEHDTFFPIYALLRDSKIEYKKTIGERHG 60
QY 61 KIVRVADVAKKKKFLGRDVEYKLIIEFHODVPALRGKIREPVAVIDIYEDIPARKY 120
DB 61 KIVRVADVAKKKKFLGRDVEYKLIIEFHODVPALRGKIREPVAVIDIYEDIPARKY 120
QY 121 LIDKGLIPMEGEELKLAFLDIETLHGEDEGCKEITIMISYADENAKVITWKMDLPY 180
DB 121 LIDKGLIPMEGEELKLAFLDIETLHGEDEGCKEITIMISYADENAKVITWKMDLPY 180
QY 181 VDVVNEREMIKRFQIVREKDPDVLITYNGDNFDLPYLKRAEKLGYTLGRDKEHE 240
DB 181 VDVVNEREMIKRFQIVREKDPDVLITYNGDNFDLPYLKRAEKLGYTLGRDKEHE 240
QY 239 PKMQIGDWTAVEVGRIFHFDLYHVTITINLPYTLVEAVIAGKREKREKVADEIAKA 298
DB 239 PKMQIGDWTAVEVGRIFHFDLYHVTITINLPYTLVEAVIAGKREKREKVADEIAKA 298
QY 301 WTEESMKKLAQYSMEDARATYELKEFFPMELAKLIGQSVMPVSSSGNLYEWYLL 360
DB 301 WTEESMKKLAQYSMEDARATYELKEFFPMELAKLIGQSVMPVSSSGNLYEWYLL 360
QY 359 WESGENSELKVAYSMEDAKATYELKEFLPMETQLSRVGOPLMDVSSSGNLYEWYLL 358
DB 359 WESGENSELKVAYSMEDAKATYELKEFLPMETQLSRVGOPLMDVSSSGNLYEWYLL 358
QY 421 VSPDTLREBCKKNYDVAIVGKFKDPGFIPLSLGELLITWRIKKMKATDPIEKK 480
DB 421 VSPDTLREBCKKNYDVAIVGKFKDPGFIPLSLGELLITWRIKKMKATDPIEKK 480
QY 499 VSPDTLREBCKKNYDVAIVGKFKDPGFIPLSLGELLITWRIKKMKATDPIEKK 478
DB 499 VSPDTLREBCKKNYDVAIVGKFKDPGFIPLSLGELLITWRIKKMKATDPIEKK 478
QY 481 MDYRQAVKTLHANSYGYGMVKARWYSKCEASVTAMGRYIMTILKEIEKRGFVL 540
DB 481 MDYRQAVKTLHANSYGYGMVKARWYSKCEASVTAMGRYIMTILKEIEKRGFVL 540
QY 479 LLDYRQKAKRLANSFYGYGAKAMWCKECSASTANGRKYIELVMWELEKEFGKVL 538
DB 479 LLDYRQKAKRLANSFYGYGAKAMWCKECSASTANGRKYIELVMWELEKEFGKVL 538
QY 541 YADTDGFYATIPGEKPEITKKKAKKFLKYINSKLPGLLEVEGEFVLRGFPFAKKRYAVI 600
DB 539 YIDTDGLVATIPGSESEELKKALEFVKYINSKLPGLLEVEGEFVLRGFPFAKKRYAVI 598
QY 601 DEGRITTRGLEVVRDSEIAKETQAKVLEAILKEDSEKAVEIVKQVVEIAKYQVPL 660
DB 599 DEEGVITRGLTEIVRDMSEIAKETQAVLEITLKHGVEEAVRIYKIOKLANYEIRP 658
QY 661 EKLVIHEOITKDSYKAGPHVAIAKRLAAGIKRPGTIISYILRSGKISDPAVL 720
DB 659 EKLAIVEOTRPLHEKKAIPHAAKLAAGVKIKPGMIGVILRGGDPLISRAILA 718
QY 721 SEYDPKHHKYPDDYIENOVLPVLRILEAFGRKEDLKQSSQOYGLDAML 772
DB 719 EYDPRKKHKYDAEYIENOVLPVLRILEGFGYRKEDLRYQKTRQVGLTSM 770

```

RESULT 11
US-07-803-627A-1

Sequence 1, Application US/07803627A

Patent No. 5948663

GENERAL INFORMATION:

APPLICANT: MATHUR, Eric J.

TITLE OF INVENTION: Purified Thermostable Pyrococcus

TITLE OF INVENTION: furiosus DNA Polymerase I

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

FACTORY AFFILIATION DATA:

APPLICATION NUMBER: US 07/686,340
FILING DATE: 17-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/626,057
FILING DATE: 11-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/513,994
FILING DATE: 26-APR-1990
ATTORNEY/AGENT INFORMATION:
NAME: WILLIAMS, GREGORY D.
REGISTRATION NUMBER: 30901
REFERENCE/DOCKET NUMBER: NEB-101
TELECOMMUNICATION INFORMATION:
TELEPHONE: (508) 927-5054
TELEFAX: (508) 927-1705
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1022 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
-08-271-364A-8

Query Match 62.2%; Score 2502.5; DB 1; Length 1022;
Best Local Similarity 62.9%; Pred. No. 1.2e-186;
Matches 534; Conservative 59; Mismatches 129; Indels 127; Gaps 17;

QY 1 MFDPDYIKKDGPIIRIKKENGKFIKLDPHOPYIYALKDSIDKIKGERRG 60
DB 1 MLDIDYITKDGPIIRIKKENGKFIKLDPHOPYIYALKDSIDKIKGERRG 60
QY 61 KIVRVDAVKKKFLGSDVEWKLIFEPHODVPALGKIREHPAVIDIYEDIPAKRY 120
DB 61 KIVRVDAVKKKFLGSDVEWKLIFEPHODVPALGKIREHPAVIDIYEDIPAKRY 120
QY 121 LIDKGLIPMEGDEELKLMFDFETFEHEDERGKELIMISYADEEARVITKNDLPY 180
DB 121 LIDKGLIPMEGDEELKLMFDFETFEHEDERGKELIMISYADEEARVITKNDLPY 180
QY 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKLGYTLIGRDKEHPE 240
DB 181 VDVVSNEREMIKRFVQIVREKDPDLITNGDNFDLPYLKRAEKLGYTLIGRDKEHPE 240
QY 241 PRIHRGDSFAVEIKGRIHFDLFPVVRTINLPTYLAVYEAVLGKTSKLGAEETAAI 300
DB 241 PRIHRGDSFAVEIKGRIHFDLFPVVRTINLPTYLAVYEAVLGKTSKLGAEETAAI 300
QY 301 WETEESSMKLAQYSMDARATYELGKEFFPMELAKLIGOSYWDVSRSTGNLVEMYLL 360
DB 301 WETEESSMKLAQYSMDARATYELGKEFFPMELAKLIGOSYWDVSRSTGNLVEMYLL 360
QY 361 RYAVENNELAPNKPDEEYRRRLRTTYLGYYKEPERGLMENTTYLDFPCLYPSIIVTNH 420
DB 361 RYAVENNELAPNKPDEEYRRRLRTTYLGYYKEPERGLMENTTYLDFPCLYPSIIVTNH 420
QY 421 VSPDTEREGCKKNYDAPITGVYKFCDFPGFIPISIGELITMRQETIKKMKATIDPIEKK 480
DB 421 VSPDTEREGCKKNYDAPITGVYKFCDFPGFIPISIGELITMRQETIKKMKATIDPIEKK 480
QY 481 MLDYRORAVKLANSTY-----GYMGYPK----- 504
DB 481 MLDYRORAVKLANSTY-----GYMGYPK----- 504
QY 505 -----ARWYKSCAAS-----VTAMGRHYIEMTIKEIEKFGFV-----LYADTDGFY 548
DB 505 -----ARWYKSCAAS-----VTAMGRHYIEMTIKEIEKFGFV-----LYADTDGFY 548
QY 541 VNNLFPFSFNKKIKSEVKKVVALIRHKKYKGAKEYIQLSSGKRIINTAGHSLETVRNGEI 600
DB 541 VNNLFPFSFNKKIKSEVKKVVALIRHKKYKGAKEYIQLSSGKRIINTAGHSLETVRNGEI 600
QY 549 ATIPGE-----KPEYIKKRAKEFLKYINSKLPGLLELEYEGFY-----RGFF- 591
DB 549 ATIPGE-----KPEYIKKRAKEFLKYINSKLPGLLELEYEGFY-----RGFF- 591
QY 601 KEVSGGCIKEGDLIVAPKKIKLNEKGVSTINIPELISDLSSEETPADIVMTISAKGRKNFER 660
DB 601 KEVSGGCIKEGDLIVAPKKIKLNEKGVSTINIPELISDLSSEETPADIVMTISAKGRKNFER 660

QY 592 --VAKKRYVIDEGRIIT-----RGLEVVRDWSSEIAKETQA-KVL 630
DB 661 GMLRTLMMMEGEENRRIIFENRYLFHLEKLGILKILPREGYEV--DWERLKKYKQLYEKL 718
QY 631 EALIKEDSVKNAVEIKDVVEELAKYQVPLEKVIHEQITKDSSEKKAIGPHVIAKKRLA 690
DB 719 AGSVKYNKMKREYLVNFNEIKDFISY-FP-----QKELEWK-IGTLNGFPFTNCI 766
QY 691 AKGIKVPCTIISYIVLRG-----SGKISDRVILLSEYDPKHKYDPDYIENQVLP 742
DB 767 LK-VDEDFGKILGYVSEGYAGAKKKTGIGS-----YSVKLYNEDPN-----VLE 811
QY 743 AVRLITLAF 751
DB 812 SMKRVAEKF 820

Search completed: May 28, 2002, 14:17:33
Job time: 31 sec